

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:16:53 ; Search time 54 Seconds
(without alignments)
905.198 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHPWFKRTILGPEFYS.....HAERAIPIVSREKXTAPSS 173

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	907	99.0	173	7	ADP75366	Adp75366 Human SHS
2	702	76.6	177	6	ABG76084	Abg76084 Human chi
3	489	53.4	175	3	AAy07074	AAy07074 Renal can
4	489	53.4	175	3	AAg03886	AAg03886 Human sec
5	489	53.4	175	7	ADP56087	ADP56087 Human pro
6	489	53.4	175	7	ADP75367	ADP75367 Human SHS
7	488	53.3	175	3	AAg03887	AAg03887 Human sec
8	488	53.3	175	2	AAAR87512	AAAR87512 Human alp
9	485	52.9	175	6	ABM04835	ABM04835 Rat alpha
10	476	52.0	175	7	ADBP9861	ADBP9861 Mouse alp
11	340	37.1	160	2	AAAR78692	AAAR78692 Human ske
12	340	37.1	160	3	AAAB42472	AAAB42472 Human ORF
13	340	37.1	160	5	ABG93941	ABG93941 Human pol
14	340	37.1	160	5	ABP43873	ABP43873 Human ske
15	340	37.1	160	7	ADP75368	ADP75368 Human SHS
16	340	37.1	165	2	AAAY60544	AAAY60544 Human nor
17	336	36.7	162	2	AAAR78693	AAAR78693 Rat skele
18	336	36.7	162	5	ABG93942	ABG93942 Rat polyp
19	336	36.7	162	7	ADP56085	ADP56085 Rat Prote
20	330	36.0	108	3	AAAG03888	AAAG03888 Human sec
21	308.5	33.7	187	4	ABB60738	ABB60738 Drosophila
22	289.5	31.6	205	5	ABBT7897	ABBT7897 Human HSB
23	289.5	31.6	205	5	AAO18753	AAO18753 Human HSP
24	289.5	31.6	205	5	ADP83437	ADP83437 Human pro
25	289.5	31.6	205	7	ADP75363	ADP75363 Human SHS

26	289.5	31.6	459	3	AAAB22936	AAAB22936 GFP-HSP27
27	289.5	31.6	459	5	ABG94499	ABG94499 Protease
28	289	31.6	206	7	ADP83435	ADP83435 Rat Prote
29	288	31.4	205	2	AAAR41022	AAAR41022 Protein h
30	288	31.4	471	2	AAW53352	AAW53352 Heat shoc
31	288	31.4	471	6	ABU04599	ABU04599 Human exp
32	284.5	31.1	102	5	AAU99185	AAU99185 Target mo
33	283	30.9	232	3	AAAB11397	AAAB11397 E. coli e
34	283	30.9	232	4	AAAB74198	AAAB74198 OmpA-Hsp2
35	283	30.9	232	4	AAAB70768	AAAB70768 Expressio
36	283	30.9	232	4	AAAY72019	AAAY72019 E. coli O
37	276	30.1	182	4	ABU52916	ABU52916 Human met
38	266	29.0	136	5	ABG93940	ABG93940 Deer poly
39	265.5	29.0	201	2	AAAR05780	AAAR05780 Protein P
40	238	26.0	182	7	ADP75365	ADP75365 Human SHS
41	214	23.4	166	4	ABB60677	ABB60677 Drosophila
42	202.5	22.1	208	4	ABB60474	ABB60474 Drosophila
43	196	21.4	213	4	ABB60679	ABB60679 Drosophila
44	193	21.1	199	4	ABB60477	ABB60477 Drosophila
45	174.5	19.1	445	4	ABB60464	ABB60464 Drosophila

ALIGNMENTS

RESULT 1

ADP75366 standard; protein; 173 AA.

ADP75366 (first entry)
29-JAN-2004 (first entry)
Human SHSP family member, alpha-crystallin.

Hybrid protein chaperone; protein stabilisation; heat shock protein;
SHSP family; protein aggregation inhibition; cell death inhibition;
genome stability pathway inhibition; protein denaturation identification;
protein conformation related disease; cardiomyopathy; cataract;
neurodegenerative disease; cardiac; ophthalmological; neuroprotective;
gene therapy; alpha-crystallin; human.

Homo sapiens.
MO2003091266-A2.
06-NOV-2003.

23-APR-2003; 2003MO-GB001721.
23-APR-2002; 2002GB-0000334.

(UYDU-) UNIT DUNDEE.

Quinlan R;

WPI; 2003-865571/80.

New hybrid protein chaperone (e.g. heat shock protein) useful for
stabilizing proteins and/or protein activities, or as an agent to prevent
protein aggregation, or for treating diseases involving altered protein
conformations.

Disclosure; Fig 11; 45pp; English.

The invention relates to a hybrid protein chaperone for stabilising
proteins and/or protein activities. Protein chaperones (also known as
heat shock proteins) are divided into 4 families on the basis of their
primary sequence and chaperone properties: HSP90, HSP70, HSP60 and SHSP
(small heat shock protein). The invention is based upon the finding that
among the SHSP family, which have a general structure of a central domain
(called the alpha-crystallin domain) flanked by N and C-terminal regions,
replacement of one or more regions of an SHSP with the corresponding

CC region from a second sHSP can improve the activity compared to native
 CC sHSPs. In a particular embodiment of the invention, the hybrid chaperone
 CC is a hybrid sHSP designated alphaB-HSP27 comprising the N-terminus and
 CC central portion of alphaB-crystallin and the C-terminal tail of HSP27.
 CC However, the hybrid protein chaperones of the invention can comprise
 CC regions from HSP90, HSP70 and HSP60 families as well as from the sHSP
 CC family. The invention also encompasses methods for stabilizing proteins
 CC such as enzymes, therapeutic proteins, diagnostic proteins, reporter
 CC proteins or antibodies, their fragments or conjugates in an aqueous
 CC solution using hybrid protein chaperones; stabilised protein formulation
 CC comprising at least one protein associated with the above hybrid protein
 CC chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
 CC and the coexpression of a recombinant protein of interest and a hybrid
 CC protein chaperone. The hybrid protein chaperones are useful as agents to
 CC prevent protein aggregation, as inhibitors of cell death and genome
 CC stability pathways, for identification of proteins that are in the
 CC process of unfolding, for the treatment of diseases involving altered
 CC protein conformations (e.g., cardiomyopathies, cataracts or
 CC neurodegenerative diseases), or for the manufacture of a medicament for
 CC the treatment of such diseases. Sequences ADE75363-ABE75373 represent
 CC members of the sHSP family which may be used in hybrid protein chaperones
 CC of the invention.

SO Sequence 173 AA;

Query March 99.0%; Score 907; DB 7; Length 173;

Best Local Similarity 99.4%; Pred. No. 4.3e-97;

Matches 172; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDVTIOHWEKRTIGPVPSPRLFDQFEGGLPEYDLAPLSTSTSPYRQSLFTVLDSG 60

DB 1 MDVTIOHWEKRTIGPVPSPRLFDQFEGGLPEYDLAPLSTSTSPYRQSLFTVLDSG 60

QY 61 ISEVSRDQKVIPLDVVHGFSPEDLTVMQDDFVEIHGKHNKRODDHGYSREFFHRRYL 120

DB 61 ISEVSRDQKVIPLDVVHGFSPEDLTVMQDDFVEIHGKHNKRODDHGYSREFFHRRYL 120

QY 121 PSNVDSALSCSLSDADGMLTFGCPKIQTGDLTAHERAIPVSREKPTAPSS 173

DB 121 PSNVDSALSCSLSDADGMLTFGCPKIQTGDLTAHERAIPVSREKPTAPSS 173

RESULT 2

ABG76084 standard; protein; 177 AA.

AC ABG76084;

DT 16-MAY-2003 (first entry)

DE Human chimeric protein alpha BNAC crystallin.

KW Human; crystallin; chimeric; alpha BNAC crystallin; protein shelf life;

KW protein aggregation; accessible hydrophobic region increase; mutant;

KW larger size oligomer formation; intersubunit interaction increase;

KW larger aggregate formation; larger porous oligomer formation;

KW increased ellipticity; less solvent accessible tryptophan; muten;

KW increased chaperone-like activity; alpha A crystallin;

KW alpha B crystallin.

XX Homo sapiens.

OS Synthetic.

XX Key

XX Location/Qualifiers

XX Region

XX 1..82

XX /note= "Alpha B crystallin N-Terminal"

XX 83..177

XX /note= "Alpha A crystallin C-Terminal"

XX US2002177192-A1.

XX 28-NOV-2002.

XX PD

XX XX

PF 26-MAR-2002; 2002US-00105427.

PR 28-MAR-2001; 2001US-0279223P.

XX (KUMA/) KUMAR L V S.

PA (RAOC/) RAO C M.

XX Kumar LVS, Rao CM;

DR WPI; 2003-298776/29.

DR N-PSDB; ABX12062.

XX New chimera alpha BNAC nucleic acid, useful for preventing aggregation of

XX proteins and also for increasing shelf life of proteins of pharmaceutical

XX value.

XX Claim 4; Fig 10; 17pp; English.

CC The invention relates to a chimera alpha BNAC polynucleotide that encodes

CC a chimeric alpha BNAC polypeptide. The polypeptide is also useful for

CC preventing protein aggregation. The polypeptide is also useful for

CC increasing the shelf life of proteins of pharmaceutical value. The

CC polypeptide shows an increase in accessible hydrophobic regions, forms

CC larger size oligomers, shows an increase in intersubunit interaction,

CC forms larger aggregates, forms larger porous oligomers and shows

CC increased ellipticity as compared to eye lens crystallins alpha A and

CC alpha B. The tryptophan residues in the polypeptide are less solvent

CC accessible as compared to those of eye lens crystallins alpha A and alpha

CC B. The polypeptide shows extraordinarily high chaperone-like activity

CC ranging between 3 - 6 times that of the eye lens crystallins alpha A and

CC alpha B. The present sequence represents the amino acid sequence of the

CC human alpha BNAC crystallin chimeric protein

CC Sequence 177 AA;

QY 1 MDVTIOHWEKRTIGPVPSPRLFDQFEGGLPEYDLAPLSTSTSPYRQSLFTVLDSG 60

DB 1 MDVTIOHWEKRTIGPVPSPRLFDQFEGGLPEYDLAPLSTSTSPYRQSLFTVLDSG 60

QY 56 VLDSGISEVSRDQKVIPLDVVHGFSPEDLTVMQDDFVEIHGKHNKRODDHGYSREFFHRRYL 115

DB 60 WFDYGLSEMRLEKDRFSVNLVDKHFSPEDLTVMQDDFVEIHGKHNKRODDHGYSREFFHRRYL 119

QY 116 RRYRLPSNVDSALSCSLSDADGMLTFGCPKIQTGDLTAHERAIPVSREKPTAPSS 173

DB 120 RRYRLPSNVDSALSCSLSDADGMLTFGCPKIQTGDLTAHERAIPVSREKPTAPSS 177

RESULT 3

AAAY07074 standard; protein; 175 AA.

AC AAAY07074;

DT 02-JUL-1999 (first entry)

DE Renal cancer associated antigen precursor sequence.

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer.

XX Homo sapiens.

XX W09904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998; 98WO-US014679.

XX PD

XX XX

[illegible]

XX Hybrid protein chaperone; protein stabilisation; heat shock protein;
 XX SHSP family; protein aggregation inhibition; cell death inhibition;
 XX genome stability pathway inhibition; protein denaturation identification;
 XX protein conformation related disease; cardiomyopathy; cataract;
 XX neurodegenerative disease; cardiac; ophthalmological; neuroprotective;
 XX gene therapy; alphan-crystallin; human.
 XX Homo sapiens.
 XX WO2003091266-A2.
 XX PD 06-NOV-2003.
 XX PF 23-APR-2003, 2003WO-GB001721.
 XX PR 23-APR-2002; 2002GB-00009334.
 XX (UVDU-) UNIV DUNDEE.
 XX Quinlan R;
 XX WPI; 2003-865571/80.
 XX New hybrid protein chaperone (e.g. heat shock protein) useful for
 XX stabilizing proteins and/or protein activities, or as an agent to prevent
 XX protein aggregation, or for treating diseases involving altered protein
 XX conformations.
 XX
 XX Disclosure; Fig 11; 45pp; English.
 XX
 XX The invention relates to a hybrid protein chaperone for stabilising
 XX proteins and/or protein activities. Protein chaperones (also known as
 XX heat shock proteins) are divided into 4 families on the basis of their
 XX primary sequence and chaperone properties: HSP90, HSP70, HSP60 and SHSP
 XX (small heat shock protein). The invention is based upon the finding that
 XX among the SHSP family, which have a general structure of a central domain
 XX (called the alpha-crystallin domain) flanked by N and C-terminal regions,
 XX replacement of one or more regions of an SHSP with the corresponding
 XX region from a second SHSP can improve the activity compared to native
 XX SHSPs. In a particular embodiment of the invention, the hybrid chaperone
 XX is a hybrid SHSP designated alphan-HSP27 comprising the N-terminus and
 XX central portion of alphan-crystallin and the C-terminal tail of HSP27.
 XX However, the hybrid protein chaperones of the invention can comprise
 XX regions from HSP90, HSP70 and HSP60 families as well as from the SHSP
 XX family. The invention also encompasses methods for stabilising proteins
 XX such as enzymes, therapeutic proteins, diagnostic proteins, reporter
 XX proteins or antibodies, their fragments or conjugates in an aqueous
 XX solution using hybrid protein chaperones; stabilised protein formulation
 XX comprising at least one protein associated with the above hybrid protein
 XX chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
 XX and the coexpression of a recombinant protein of interest and a hybrid
 XX protein chaperone. The hybrid protein chaperones are useful as agents to
 XX prevent protein aggregation, as inhibitors of cell death and genome
 XX stability pathways, for identification of proteins that are in the
 XX process of unfolding, for the treatment of diseases involving altered
 XX protein conformations (e.g., cardiomyopathies, cataracts or
 XX neurodegenerative disease), or for the manufacture of a medicament for
 XX the treatment of such diseases. Sequences ADE75363-ADE75373 represent
 XX members of the SHSP family which may be used in hybrid protein chaperones
 XX of the invention.
 XX
 XX Sequence 175 AA;
 XX
 XX Query Match 53.4%; Score 489; DB 7; Length 175;
 XX Best Local Similarity 54.5%; Pred. No. 2,3e-48;
 XX Matches 97; Conservative 34; Mismatches 33; Indels 12; Gaps 6

Db 60 WPDGISEMERLEKRFVNDVKGHPSPEELKVKVLGVIVHGHGERODEHGFIISREFH 119
 QY 116 RRYRLPSPNVDSALSCSLSDAGMLTFCGPKIQTGLDTHAERAIIPVREKRP--TSAP 171
 120 RKYRIPADVDPLRTITSSLSGDLTVNGPRKQ-----VSGERTIPITREKPAVTAP 173

RESULT 7
 AAG03887
 ID AAG03887 standard; protein; 250 AA.

XX AAG03887;
 AC AAG03887;
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein, SEQ ID NO: 7968.
 XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping.
 XX Homo sapiens.
 XX EP1033401-A2.
 PN 06-SEP-2000.
 PD 21-FEB-2000; 2000EP-00200610.
 PF 26-FEB-1999; 99US-0122487P.
 PR (GEST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI: 2000-500381/45.
 DR N-PSDB; AAC03893.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 PS Claim 13; SEQ ID NO 7968; 71pp + Sequence Listing; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number of
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
 CC sequences derived from the 5' ends of mRNAs and even in those cases where
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
 CC are also used in diagnostic, forensic, gene therapy and chromosome
 CC mapping procedures. They are used to obtain upstream regulatory sequences
 CC and to design expression and secretion vectors
 CC
 XX
 SQ Sequence 250 AA;

Query Match 53.4%; Score 489; DB 3; Length 250;
 Best Local Similarity 54.5%; Pred. No. 3.9e-48;
 Matches 97; Conservative 34; Mismatches 35; Indels 12; Gaps 6;

QY 1 MDVTIHPWFKRTLGFFY-PSRLFDQFSGLFYDLPPLSLSTISPYRQ--SLR--T 55
 Db 76 MDALHHPWIRKRPFPSPSRFLDQFGBHLSDLPF-TSTSLSPFYIRPSPFRAAP 134
 QY 56 VLDSGISEVRSDDKRVIFLDVKGHPSPEELKVKVLGVIVHGHGERODEHGFIISREFH 115
 Db 135 WPDGISEMERLEKRFVNDVKGHPSPEELKVKVLGVIVHGHGERODEHGFIISREFH 194
 QY 116 RRYRLPSPNVDSALSCSLSDAGMLTFCGPKIQTGLDTHAERAIIPVREKRP--TSAP 171

Db 195 RKYRIPADVDPLRTITSSLSGDLTVNGPRKQ-----VSGERTIPITREKPAVTAP 248
 RESULT 8
 AAR87512
 ID AAR87512 standard; protein; 175 AA.

XX AAR87512;
 AC AAR87512;
 DT 10-JUL-1996 (first entry)
 DE Human alpha B crystallin.
 XX
 KM Human; alpha B crystallin; eye lens; striated muscle; kidney; CNS tissue;
 KM central nervous system; oligodendrocyte; white matter;
 KM heat shock protein; chaperone characteristic; stress;
 KM Alzheimer's disease; Alexander's disease; antibody; multiple sclerosis;
 KM MS; lesion; autoantigen; bovine.
 XX
 OS Homo sapiens.
 XX
 PN W09533997-A1.
 XX
 PD 14-DEC-1995.
 XX
 PF 08-JUN-1995; 95WO-NL000203.
 XX
 PR 09-JUN-1994; 94EP-00201653.
 XX
 PA (NEDE) NEDERLANDSE ORG TNO.
 PI Van Noort JM, Van Sechel AC, Ouagmiri ME;
 XX
 DR WPI; 1996-040357/04.
 XX
 PT Use of alpha B crystallin in diagnosis and treatment of auto-immune
 PT disease - esp. multiple sclerosis, also therapeutic use of specific
 PT antagonists, auto-reactive T cells or their T-cell receptors.
 XX
 PS Example 2; Fig 3; 37pp; English.
 XX
 CC This sequence represents human alpha B crystallin. This protein os
 CC located in the eye lens, and also in striated muscle, kidney and CNS
 CC tissue. Within the normal CNS, alpha B crystallin is mainly found in
 CC oligodendrocytes located in the white matter. It has been shown that
 CC alpha B crystallin is highly homologous to heat shock proteins and has
 CC been found to exhibit chaperone characteristics. Immunohistochemical
 CC studies have shown that cellular levels of alpha B crystallin increase as
 CC a result of stress. Also in a number of pathological conditions, e.g.
 CC Alzheimer's and Alexander's disease, levels of alpha B crystallin are
 CC elevated. Immunohistochemical staining of CNS tissues with anti-alpha B
 CC crystallin antibodies revealed enhanced expression within or close by
 CC multiple sclerosis (MS) lesions as compared to unaffected regions of
 CC white matter. Alpha B crystallin is thought to be the autoantigen in MS
 CC and is the target of various modifications, e.g. phosphorylation.
 CC There are only four amino acid differences between bovine and human alpha
 CC B crystallin
 CC
 XX
 SQ Sequence 175 AA;

Query Match 53.3%; Score 488; DB 2; Length 175;
 Best Local Similarity 54.5%; Pred. No. 3e-48;
 Matches 97; Conservative 35; Mismatches 34; Indels 12; Gaps 6;

QY 1 MDVTIHPWFKRTLGFFY-PSRLFDQFSGLFYDLPPLSLSTISPYRQ--SLR--T 55
 Db 1 MDALHHPWIRKRPFPSPSRFLDQFGBHLSDLPF-TSTSLSPFYIRPSPFRAAP 59
 QY 56 VLDSGISEVRSDDKRVIFLDVKGHPSPEELKVKVLGVIVHGHGERODEHGFIISREFH 115
 Db 60 WPDGISEMERLEKRFVNDVKGHPSPEELKVKVLGVIVHGHGERODEHGFIISREFH 119


```

XX KM p20; stress protein, muscle; diagnosis, autoimmune disease, dimer.
XX OS Homo sapiens.
XX PN JP07181180-A.
XX PD 21-JUL-1995.
XX PF 24-DEC-1993; 93JP-00327666.
XX PR 24-DEC-1993; 93JP-00327666.
XX PA (KATO/) KATO K.
XX PG (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.
XX DR WPI; 1995-285570/38.
XX PT New stress protein p20, its purification and synthesis, and an antibody
PT directed against it - useful as a diagnostic reagent in the detection of
PT auto-immune disease.
XX PS Claim 1; Page 2; 9pp; Japanese.
XX CC The present sequence is that of a new human stress protein p20. The new
CC stress protein can be used as a diagnostic agent for autoimmune disease.
CC Antibodies raised against the protein can be used in immunoassays to
CC detect the presence of the protein and in a new protein purification
CC process
XX SQ Sequence 160 AA;

Query Match          37.1%; Score 340; DB 2; Length 160;
Best Local Similarity 40.1%; Pred. No. 4.8e-31;
Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

QY 3 VTIDHPWFKRTLGPF---YPSRLFDQFPGSGLFYDILPLSLTISPIYRQSLFRVLVD 58
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 VPVQPSWLRRAASAPLPGLSAPGRFLDQRFSGLEAEIALALCPPTLAPY---LRAPSV 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 59 SGISEVSDSDRDKFYIFDVKHSPEDELTVKQDDFVEIHGKNRRODGHGYSREPHRY 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 LPVAQVTDGHRFSVLVDVKNHFSPEELAVKVGSHVEYHARHERPDHGVAREFHRY 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 RLPSNVDSALSCSLSDAGMLTFQGPRIQTGLDATHAERALPVSRREKPTSA 170
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 122 RLPPGVDPAAVTSALSPGVLST-----QAAPAAQAAPPAA 158

RESULT 12
AAB42472
ID AAB42472 standard; protein, 160 AA.
XX
AC AAB42472;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human OREF ORF236 polypeptide sequence SEQ ID NO:4472.
XX
KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypertoid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

```

```

XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US008621.
XX PR 31-MAR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 05-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinketsu RA, Leach M;
XX DR WPI; 2000-602362/57.
XX DR N-PSDB; AAC76681.
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX neurodegenerative disorders and cardiovascular disease.
XX PS Claim 1; Page 3657; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human OREF open reading frames 1 to 3161. The OREF
XX sequences have activities such as: cytosolic; hepatotropic; vulnerable;
XX antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antineumatic; antihypertoid; and antianaemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX OREF-associated disorder. The nucleic acids can be used to express OREF
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX SQ Sequence 160 AA;

Query Match          37.1%; Score 340; DB 3; Length 160;
Best Local Similarity 40.1%; Pred. No. 4.8e-31;
Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

QY 3 VTIDHPWFKRTLGPF---YPSRLFDQFPGSGLFYDILPLSLTISPIYRQSLFRVLVD 58
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 VPVQPSWLRRAASAPLPGLSAPGRFLDQRFSGLEAEIALALCPPTLAPY---LRAPSV 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 59 SGISEVSDSDRDKFYIFDVKHSPEDELTVKQDDFVEIHGKNRRODGHGYSREPHRY 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 LPVAQVTDGHRFSVLVDVKNHFSPEELAVKVGSHVEYHARHERPDHGVAREFHRY 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 RLPSNVDSALSCSLSDAGMLTFQGPRIQTGLDATHAERALPVSRREKPTSA 170
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 122 RLPPGVDPAAVTSALSPGVLST-----QAAPAAQAAPPAA 158

RESULT 13
ABG93941
ID ABG93941 standard; protein, 160 AA.
XX
AC ABG93941;
XX

```

DT 26-NOV-2002 (first entry)

XX Human polypeptide orthologous to DACC-9.

DE Human polypeptide orthologous to DACC-9.

XX Human; deer; rat; mouse; DACC; deer antler cartilage cell;

KM cell stimulation; cell inhibition; cell growth; cell division;

KM mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;

KM repair; regeneration; restoration; extracellular matrix;

KM cartilaginous matrix; cartilage; disc; connective tissue; agonist;

KM antagonist; gene therapy.

XX Homo sapiens.

XX WO200264625-A1.

XX PD 22-AUG-2002.

XX 15-FEB-2002; 2002WO-AU000163.

XX 15-FEB-2001; 2001AU-00003116.

XX (ADPP-) ADP PHARM PTY LTD.

XX (UNSY) UNIV SYDNEY.

XX PA Roubin R, Ghosh P;

XX WPI; 2002-643456/69.

XX Stimulating or inhibiting cell growth and/or division, useful for

PT stimulating chondrogenesis, cartilage, disc or connective tissue growth,

PT repair, and/or regeneration, comprises administering deer antler

PT cartilage gene.

XX Claim 11; Page 136-137; 214pp; English.

XX The invention discloses a method for stimulating or inhibiting cell

CC growth and/or division which comprises contacting or inserting into an

CC animal cell a polypeptide comprising one of the deer antler cartilage

CC cell (DACC) clones disclosed. More particularly, the method relates to

CC these polypeptides stimulating mesenchymal cell growth and/or division

CC and to transfecting these cells and chondrocytes with vectors carrying

CC the genes of these polypeptides capable of stimulating chondrogenesis,

CC osteogenesis, growth, repair, regeneration and/or restoration of the

CC extracellular matrix. The chondrocytes selectively express genes required

CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides

CC are useful for identifying an agent that modulates the activity of the

CC polypeptide, for stimulating mesenchymal cell growth and/or division by

CC exposing animal mesenchymal cells to conditioned media or its active

CC fraction, obtained from deer antler cartilage cells, for inhibiting cell

CC growth and/or division by inserting into an animal cell, a compound which

CC inhibits the translation of the polynucleotide encoding the DACC. The

CC method and the translation of the polynucleotide encoding the DACC. The

CC growth and/or division or for stimulating chondrogenesis, cartilage, disc

CC or connective tissue growth, repair, regeneration and/or restoration in

CC an animal. The polynucleotides, polypeptides, agonists and antagonists

CC may be used in treatment modalities, specifically in gene therapy. The

CC polypeptides can be used as bait proteins in a two- or three-hybrid assay

CC to identify other proteins, which bind to or interact with the

CC polypeptide and are involved in modulating cell growth and/or division.

CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by

CC the DACC cDNA clones

XX Sequence 160 AA;

XX

Query Match 37.1%; Score 340; DB 5; Length 160;

Best Local Similarity 40.1%; Pred. No. 4.8e-31;

Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

OY 3 VTIOHPWFKRTLGPF---YPSRLFDQFFGGLFEYDILLPLSLSTISFYQSFLRTYLD 58

DB 5 VPVQPSWLRASAPLPGLSAPGRFLFDQRFSGGLAEALALCPPTLAIY---LEAPSYA 61

OY 59 SGISEVRSRDRKVFILVDVKHPSPEDLLVXQDDVDEIHGKNEQDDHGYISREFHRY 118

DB 62 LPVAQVPDPCGHFVILDVKHFSPERIEAVKVGHEVHAHBEERPDHGFVAREFHRY 121

OY 119 RLPSNVDSALSCSLADGMLTFGSPKIQTGIDAHARAIYPSREKPTSA 170

DB 122 RLPSVDPAAVTASLSFEGVLSI-----QAAPASQAPPPAA 158

RESULT 14

ABP43873

ID ABP43873 standard; protein; 160 AA.

XX AC ABP43873;

XX DT 26-FEB-2003 (first entry)

XX DE Human skeletal muscle stress protein p20.

XX KM Neuroprotective; immunomodulator; cancer; chromosome 19pter-q12;

XX KM cytosolic; anti-inflammatory; gene therapy; nutritional supplement;

XX KM wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

XX KM amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

XX KM vulnery.

XX OS Homo sapiens.

XX WO200231111-A2.

XX PD 18-APR-2002.

XX PF 11-OCT-2001; 2001WO-US027760.

XX PR 12-OCT-2000; 2000US-00687527.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Wehrman T, Dymnac RT;

XX WPI; 2002-426278/45.

XX DR N-PSDB; ABQ61117.

XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation.

XX Claim 20; SEQ ID # 776; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The

CC activity of polynucleotides of the invention may be described as,

CC vulnerary, neuroprotective, immunomodulator, cytosolic and anti-

CC inflammatory. Compositions comprising nucleic acids of the invention are

CC useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records ABP43544-

CC ABP43989 represent polypeptides encoded by polynucleotides of the

CC invention. NOTE: The sequence data for this patent did not form part of

CC the printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 160 AA;

XX

Query Match 37.1%; Score 340; DB 5; Length 160;

Best Local Similarity 40.1%; Pred. No. 4.8e-31;

Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

OY 3 VTIOHPWFKRTLGPF---YPSRLFDQFFGGLFEYDILLPLSLSTISFYQSFLRTYLD 58

DB 5 VPVQPSWLRASAPLPGLSAPGRFLFDQRFSGGLAEALALCPPTLAIY---LEAPSYA 61

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:17:08 ; Search time 12 Seconds

(without alignments)
750.678 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916
Sequence: 1 MDVTIQHWPVKRTLGPPFPS.....HARRAPVSRREKPTAPSS 173

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	916	100.0	173 1	CRAA_HUMAN
2	884.5	96.6	172 1	CRAA_MACMU
3	884	96.5	173 1	CRAA_LOXAF
4	879	96.0	173 1	CRAA_RABIT
5	878	95.9	173 1	CRAA_CAVO
6	876	95.6	173 1	CRAA_BULFU
7	873	95.3	173 1	CRAA_OCHPR
8	873	95.3	173 1	CRAA_PPRPO
9	873	95.3	173 1	CRAA_PROCA
10	871	95.1	173 1	CRAA_MOUSE
11	869	94.9	173 1	CRAA_ARIVA
12	868	94.9	173 1	CRAA_BOVIN
13	868	94.8	173 1	CRAA_TAPIN
14	864	94.3	173 1	CRAA_BALAC
15	864	94.3	173 1	CRAA_GIRCA
16	863	94.2	173 1	CRAA_PTEPO
17	862	94.1	173 1	CRAA_CERSI
18	860	93.9	173 1	CRAA_PHOOP
19	859	93.8	173 1	CRAA_PIG
20	857	93.6	173 1	CRAA_CAMDR
21	856	93.4	173 1	CRAA_CANPA
22	856	93.4	173 1	CRAA_HORSE
23	854	93.2	173 1	CRAA_URSUR
24	850	92.8	173 1	CRAA_ORYAF
25	849.5	92.7	173 1	CRA2_MESAU
26	849.5	92.7	173 1	CRA2_MOISE
27	849.5	92.7	173 1	CRA2_RAT
28	844	92.1	173 1	CRA2_ZALCA
29	843	92.0	173 1	CRAA_MUSVI
30	839	91.6	173 1	CRAA_MANJA
31	831	90.7	173 1	CRAA_ERIEU
32	831	90.7	173 1	CRAA_MACRU
33	829.5	90.6	170 1	CRAA_CHOHO

34	828	90.4	173 1	CRAA_SPAEH	OC4211 spalax leuc
35	816.5	89.1	170 1	CRAA_TAMME	P02485 tamandua me
36	816	89.1	173 1	CRAA_DIDMA	P02503 didelphis m
37	806.5	88.0	196 1	CRA2_SPAEH	P15990 spalax leuc
38	803.5	87.7	170 1	CRAA_BRAVA	P02487 brachipus va
39	798	87.1	173 1	CRAA_TUOTE	P02506 tupiambis
40	791	86.4	173 1	CRAA_RHEAM	P02505 rhea americ
41	787	85.9	173 1	CRAA_CHICK	P02504 gallus gall
42	785	85.7	173 1	CRAA_RANCA	O91311 rana catesb
43	780	85.2	173 1	CRAA_ALIMI	P06904 alligator m
44	760	83.0	161 1	CRAA_TRIN	P02500 trichechus
45	753	82.2	167 1	CRAA_RANES	P02507 rana escul

ALIGNMENTS

RESULT 1

ID	CRAA_HUMAN	STANDARD;	PRT;	173 AA.
AC	P02489;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Alpha crystallin A chain.			
GN	CRYAA OR CRYA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	PRELIMINARY SEQUENCE.			
RX	MEDLINE=76187952; PubMed=817940;			
RA	de Jong W.W., Terwindt E.C., Bloemendaal H.;			
RT	"The amino acid sequence of the A chain of human alpha-crystallin.";			
RL	FEBS Lett. 56:310-313(1975).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
RX	MEDLINE=96139023; PubMed=8587135;			
RA	Jaworski C.J.;			
RT	"A reassessment of mammalian alpha A-crystallin sequences using DNA sequencing: implications for anthropoid affinities of tarsier.";			
RL	J. Mol. Evol. 41:901-908(1995).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
RX	MEDLINE=97112991; PubMed=8943244;			
RA	Andley U.P., Mathur S., Griest T.A., Petrasch J.M.;			
RT	"Cloning, expression, and chaperone-like activity of human alpha-crystallin.";			
RL	J. Biol. Chem. 271:31973-31980(1996).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20289799; PubMed=10830953;			
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,			
RA	Park H.-S., Toyoda A., Ishii K., Tokoi Y., Choi D.-K., Groner Y.,			
RA	Soda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,			
RA	Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,			
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,			
RA	Rosenthal A., Kudoh U., Shibuya K., Kawasaki K., Asakawa S.,			
RA	Shintani A., Saeki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,			
RA	Mitsushima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,			
RA	Scharfe M., Schen O., Desario A., Reichelt J., Kauer G., Bloeker H.,			
RA	Ransner J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,			
RA	Weinmayer S., Borzym K., Gardiner K., Nizetic D., Francis F.,			
RA	Lehrach H., Reinhardt K., Yaspo M.-L.;			
RT	"The DNA sequence of human chromosome 21.";			
RL	Nature 405:311-319(2000).			
RN	[5]			
RP	SEQUENCE OF 1-104 FROM N.A.			
RX	MEDLINE=89143747; PubMed=2918909;			
RA	Jaworski C.J., Flatisorsky U.;			

RT "A pseudo-exon in the functional human alpha A-crystallin gene.",
 RL Nature 337:752-754 (1989).
 RN [6]
 RP SEQUENCE OF 1-63 AND 166-173 FROM N.A.
 RC TISSUE=Spleen;
 RX MEDLINE=87005033; PubMed=3758227;
 RA McDevitt D.S., Hawkins J.W., Jaworski C.J., Piatigorsky J.;
 RT "Isolation and partial characterization of the human alpha
 A-crystallin gene.";
 RL Eye Res. 43:285-291 (1986).
 RN [7]
 RP SEQUENCE OF 13-21 AND 79-88
 RX MEDLINE=97152993; PubMed=8999933;
 RA Lampi K.J., Ma Z., Sih M., Shearer T.R., Smith J.B., Smith D.L.,
 RA David L.L.;
 RT "Sequence analysis of betaA3, betaA3, and betaA4 crystallins
 RT completes the identification of the major proteins in young human
 RT lens.";
 RL J. Biol. Chem. 272:2268-2275 (1997).
 RN [8]
 RP STRUCTURE OF CARBOHYDRATE.
 RX MEDLINE=92112709; PubMed=1730617;
 RA Roquemore E.P., Dell A., Morris H.R., Panico M., Reason A.J.,
 RA Savoy L.-A., Mifrow G.J., Zigler J.S. Jr., Earles B.J., Hart G.W.;
 RT "Vertebrate lens alpha-crystallins are modified by O-linked N-
 RT acetylglucosamine.";
 RL J. Biol. Chem. 267:555-563 (1992).
 RN [9]
 RP DEAMIDATION.
 RX MEDLINE=98205214; PubMed=9543632;
 RA Takemoto L.J.;
 RT "Quantitation of asparagine-101 deamidation from alpha-A crystallin
 RT during aging of the human lens.";
 RL Curr. Eye Res. 17:247-250 (1998).
 RN [10]
 RP VARIANT ZONULAR CENTRAL NUCLEAR CATARACT CYS-116.
 RX MEDLINE=98133928; PubMed=9467006;
 RA Litt M., Kramer F., la Moticella D.M., Murphy W., Lovrien E.W.,
 RA Weleber R.G.;
 RT "Autosomal dominant congenital cataract associated with a missense
 RT mutation in the human alpha crystallin gene CRYAA.";
 RL Hum. Mol. Genet. 7:471-474 (1998).
 RN [11]
 RP CHARACTERIZATION OF VARIANT ZONULAR CENTRAL NUCLEAR CATARACT CYS-116.
 RX MEDLINE=20573572; PubMed=11123904;
 RA Cobb B.A., Petrash J.M.;
 RT "Structural and functional changes in the alpha A-crystallin R116C
 RT mutant in hereditary cataracts.";
 RL Biochemistry 39:15791-15798 (2000).
 CC -1- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -1- PTM: O-glycosylated; contains N-acetylglucosamine side chains.
 CC -1- PTM: IN LENS, MOST DEAMIDATION OF ASN-101 OCCURS DURING THE FIRST
 CC 30 YEARS OF AGE, FOLLOWED BY A SMALL ADDITIONAL AMOUNT OF
 CC DEAMIDATION (APPROXIMATELY 5%) DURING THE NEXT APPROXIMATELY 38
 CC YEARS, RESULTING IN A MAXIMUM OF APPROXIMATELY 50% DEAMIDATION
 CC DURING THE LIFETIME OF THE INDIVIDUAL.
 CC -1- DISEASE: Defects in CRYAA are the cause of zonal central nuclear
 CC cataract [MIM:123580]. It is one of a considerable number of
 CC phenotypically and genetically distinct forms of autosomal
 CC dominant cataract [MIM:604219]. This congenital cataract is a
 CC common major abnormality of the eye that frequently cause
 CC blindness in infants.
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U05569; AAA97523.1; -;
 DR EMBL; U6584; AAC5090.1; -;
 DR EMBL; X14789; CAA32891.1; -;
 DR EMBL; M35628; AAA52106.1; -;
 DR EMBL; M35629; AAA52105.1; -;
 DR EMBL; AF001748; BAA95535.1; -;
 DR PIR; S03344; CYH0A.
 DR GlysultideB; P02489; -;
 DR SWISS-2DPAGE; P02489; HUMAN.
 DR Genew; HGNC:2388; CRYAA.
 DR MIM; 123580; -;
 DR MIM; 604219; -;
 DR GO; GO:0007601; P:vision; TAS.
 DR InterPro; IPR001436; Crystallin_alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; Crystallin; 1.
 DR Pfam; PF00011; HSP20; 1.
 DR PRINTS; PR00299; ACRYSTALIN.
 DR Prodom; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 KW Eye lens protein; Acetylation; Glycoprotein; Disease mutation;
 KW Vision.
 FT MOD_RES 1 1 ACETYLATION (PARTIAL).
 FT MOD_RES 101 101 DEAMIDATION (PARTIAL).
 FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT VARIANT 116 116 R -> C (in zonal central nuclear
 FT cataract; reduced chaperone-like activity
 FT and increased membrane-binding capacity).
 FT FTID=VAR_003819.
 FT S -> T (IN REF. 6).
 FT T -> H (IN REF. 2).
 SQ SEQUENCE 173 AA; 19909 MW; 81804A8439837D50 CRC64;
 Query Match 100.0%; Score 916; DB 1; Length 173;
 Best Local Similarity 100.0%; Pred. No. 6,1e-82;
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDVTIQHFWFKRTGPFYPSRLFDQFEGGLFEYDLPLFLSTTSPYRQSLFRVLDSG 60
 DB 1 MDVTIQHFWFKRTGPFYPSRLFDQFEGGLFEYDLPLFLSTTSPYRQSLFRVLDSG 60
 QY 61 ISEVRSDPKRVIFLDVGHFSPEDLTIVKODDFVRIHGKNERQDDGYISREFFRRYL 120
 DB 61 ISEVRSDPKRVIFLDVGHFSPEDLTIVKODDFVRIHGKNERQDDGYISREFFRRYL 120
 QY 121 PSNVDSALSCSLSDGMLTFGPKIQTGDPATHAERAIPIVREKPTSPSS 173
 DB 121 PSNVDSALSCSLSDGMLTFGPKIQTGDPATHAERAIPIVREKPTSPSS 173
 RESULT 2
 CAA_MACMU STANDARD; PRT; 172 AA.
 AC P02488;
 DT 21-0UT-1986 (Rel. 01, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRYAA.
 OS Macaca mulatta (rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE.
 RA de Jong W.M., van der Ouderaa F.J., Versteeg M., Groenewoud G.,
 RA van Amelsvoort J.M., Bloemendaal H.;
 RT "Primary structures of the alpha-crystallin A chains of seven
 RT mammalian species.";

```

RL  Eur. J. Biochem. 53:237-242(1975).
RN
RP  [2]
RX  MEDLINE=96139023; PubMed=8587135;
RA  Jaworski C.J.;
RT  "A reassessment of mammalian alpha A-crystallin sequences using DNA
RL  sequencing: implications for anthropoid affinities of tarsier.";
RN  J. Mol. Evol. 41:901-908(1995).
[3]
RP  CARBOHYDRATE-LINKAGE SITE SER-168.
RX  MEDLINE=96241614; PubMed=8639509;
RA  Roguemore E.P., Hart G.W.;
RL  Unpublished results, cited by:
RL  Roguemore E.P., Chevrier M.R., Cotter R.J., Hart G.W.;
RX  Biochemistry 35:3578-3586(1996).
CC  -1- FUNCTION: May contribute to the transparency and refractive index
CC  of the lens.
CC  -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC  family.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL; U24061; AAA97563.1; -.
CC  PIR; A02890; CWM0AA.
CC
DR  GlycoSuiteDB; P02486; -.
DR  InterPro; IPR001436; Crystallin alpha.
DR  InterPro; IPR003090; Crystallin_N.
DR  InterPro; IPR002068; HSP20.
DR  InterPro; IPR008978; HSP20_chap.
DR  Pfam; PF00525; crystallin; 1.
DR  Pfam; PF00011; HSP20; 1.
DR  PRINTS; PR00299; ACRYSTALLIN.
DR  ProDom; PD001193; Crystallin_N; 1.
DR  PROSITE; PS01031; HSP20; 1.
KM  Eye lens protein; Acetylation; Glycoprotein.
FT  MOD_RES 1 1 ACETYLATION (PROBABLE).
FT  CARBOHYD 168 168 O-LINKED (GLCNAC).
FT  CONFLICT 153 154 TH -> HT (IN REF. 1).
SQ  SEQUENCE 172 AA; 19792 MW; 1F7AF9066BEBB2D7 CRC64;

Query Match 96.6%; Score 884.5; DB 1; Length 172;
Best Local Similarity 97.7%; Pred. No. 7e-79;
Matches 169; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 MDVTIQHPWFKRLTGPYPRLFDQFGEGLFYDLPLFLSTTSPYRQSLFTVLDG 60
DB 1 MDVTIQHPWFKRLTGPYPRLFDQFGEGLFYDLPLFLSTTSPYRQSLFTVLDG 60
QY 61 ISEVRSRDRKFVFLDYKHSPEDLTVKVDDEVEIHGKNERODDHGYISREHRRYRL 120
DB 61 ISEVRSRDRKFVFLDYKHSPEDLTVKVDDEVEIHGKNERODDHGYISREHRRYRL 120
QY 121 PSNVDSALSCSLSDGMLTFGCPKIQTGIDATHAERAIIVSRREKTSAPSS 173
DB 121 PSNVDSALSCSLSDGMLTFGCPKIQTGIDATHAERAIIVSRREKTSAPSS 173
ID CRAA_TOXAF STANDARD; PRT; 173 AA.
AC P02496;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha crystallin A chain.
GN CRYAA.
OS Loxodonta africana (African elephant).

```

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Proboscidea; Elephantidae; Loxodonta.
CC NCBI_TaxID=9785;
CC [1]
CC SEQUENCE.
CC MEDLINE=77158093; PubMed=870070;
CC de Jong W.W., Nuy-Terwindt E.C., Versteeg M.;
CC "Primary structures of alpha-crystallin A chains of elephant, whale,
CC hyrax and rhinoceros.";
CC Biochim. Biophys. Acta 491:573-580(1977).
CC [2]
CC SEQUENCE OF 121-163 FROM N.A.
CC MEDLINE=96139023; PubMed=8587135;
CC Jaworski C.J.;
CC "A reassessment of mammalian alpha A-crystallin sequences using DNA
CC sequencing: implications for anthropoid affinities of tarsier.";
CC J. Mol. Evol. 41:901-908(1995).
CC -1- FUNCTION: May contribute to the transparency and refractive index
CC of the lens.
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U24060; AAA97562.1; -.
CC PIR; A02901; CYELAA.
CC
DR  InterPro; IPR001436; Crystallin alpha.
DR  InterPro; IPR003090; Crystallin_N.
DR  InterPro; IPR002068; HSP20.
DR  InterPro; IPR008978; HSP20_chap.
DR  Pfam; PF00525; crystallin; 1.
DR  Pfam; PF00011; HSP20; 1.
DR  PRINTS; PR00299; ACRYSTALLIN.
DR  ProDom; PD001193; Crystallin_N; 1.
DR  PROSITE; PS01031; HSP20; 1.
KM  Eye lens protein; Acetylation; Glycoprotein.
FT  MOD_RES 1 1 ACETYLATION.
FT  CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ  SEQUENCE 173 AA; 19837 MW; E6DECD8C838FC13 CRC64;

Query Match 96.5%; Score 884; DB 1; Length 173;
Best Local Similarity 95.4%; Pred. No. 7.8e-79;
Matches 165; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRLTGPYPRLFDQFGEGLFYDLPLFLSTTSPYRQSLFTVLDG 60
DB 1 MDVTIQHPWFKRLTGPYPRLFDQFGEGLFYDLPLFLSTTSPYRQSLFTVLDG 60
QY 61 ISEVRSRDRKFVFLDYKHSPEDLTVKVDDEVEIHGKNERODDHGYISREHRRYRL 120
DB 61 ISEVRSRDRKFVFLDYKHSPEDLTVKVDDEVEIHGKNERODDHGYISREHRRYRL 120
QY 121 PSNVDSALSCSLSDGMLTFGCPKIQTGIDATHAERAIIVSRREKTSAPSS 173
DB 121 PSNVDSALSCSLSDGMLTFGCPKIQTGIDATHAERAIIVSRREKTSAPSS 173
ID CRAA_RABIT STANDARD; PRT; 173 AA.
AC P02493;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha crystallin A chain.
GN CRYAA.
OS Oryctolagus cuniculus (Rabbit).

```

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RA Kraus Z., Auguste Y.N., Quinl A.N., Redd A.N., Russell S., Graw J.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE.
RC TISSUE=Lens;
RA MEDLINE=93151974; PubMed=8427639;
RA Parveen R., Smith J.B., Sun Y., Smith D.L.;
RT "Primary structures of the alpha-crystallin A chains of seven
  mammalian species."
  Eur. J. Biochem. 53:237-242(1975).
  -1- FUNCTION: May contribute to the transparency and refractive index
    of the lens.
  -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
    family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC
CC EMBL; X93582; CAA64668.1; -.
CC PIR; A02896; CYRBA.
DR InterPro; IPR001436; Crystallin_alpha.
DR InterPro; IPR003090; Crystallin_N.
DR InterPro; IPR002068; Hsp20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin; 1.
DR Pfam; PF00011; HSP20; 1.
DR PRINTS; PR00299; ACRYSTRLLIN.
DR ProDom; PD001193; Crystallin_N; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Eye lens protein; Acetylation; Phosphorylation; Glycoprotein.
FT MOD_RES 1 122 ACETYLATION.
FT CARBOHD 162 162 PHOSPHORYLATION.
FT CONFLICT 170 170 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 173 AA; 19865 MW; C12C25298CC327EA CRC64;
Query Match 96.0%; Score 879; DB 1; Length 173;
Best Local Similarity 95.4%; Pred. No.2.4e-78;
Matches 165; Conservative 5; Mismatches 3; Indels 0; Gaps 0
QY 1 MDVTIQHWPFKRTIGPPYPSRLFDQFQBGKFEYDLPFLSTTSPYYRQSLFRTVDSG 60
DB 1 MDVTIQHWPFKRTIGPPYPSRLFDQFQBGKFEYDLPFLSTTSPYYRQSLFRTVDSG 60
QY 61 ISEVRSDPKRVIPLDVGHFSPEDLTIVKQDDPVATIGKKNHRQDDHGYSSEFRRRL 120
DB 61 ISEVRSDPKRVIPLDVGHFSPEDLTIVKQDDPVATIGKKNHRQDDHGYSSEFRRRL 120
QY 121 PSNDQSLSCSLSDGMLTFCGPRTIOGLDATHERAIPVSRREKPTAPSS 173
DB 121 PSNDQSLSCSLSDGMLTFCGPRTIOGLDATHERAIPVSRREKPTAPSS 173

```

```

AC P02491;1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT Alpha crystallin A chain.
GN CRYA.
OS Cavia porcellus (Guinea pig), and
OS Pedetes capensis (Springhaas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Hystriognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141, 10023;
RN [1]
RP PARTIAL SEQUENCE.
RA de Jong W.W., Zweers A., Goodman M.;
RT "Trends in the molecular evolution of alpha-crystallin.";
RL (In) Feereers H. (eds.);
RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
RL Pergamon Press, Oxford (1980).
CC -1- FUNCTION: May contribute to the transparency and refractive index
CC of the lens.
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC family.
CC PIR, A02894; CYCPAA.
DR PIR, E94432; CYCPAA.
DR InterPro: IPR001436; Crystallin alpha.
DR InterPro: IPR003090; Crystallin_N.
DR InterPro: IPR002068; HSP20.
DR InterPro: IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin; 1.
DR PRINTS; PR00129; ACRYSTALIN.
DR ProDom; PD001193; Crystallin_N; 1.
DR PROSITE; PS01031; HSP20; 1.
DR Eye lens protein; Acetylation; Glycoprotein.
FT MOD_RS 1 ACETYLATION (PROBABLE).
FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 173 AA; 19807 MW; 6D8B3E147B3A5C CRC64;

Query Match 95.9%; Score 878; DR 1; Length 173;
Best Local Similarity 95.4%; Pred. No. 3e-78;
Matches 165; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHPMFARLTGPPYPSRLFPQFGEGLFXYDLPLFLSSTISPPYRQSLFRTVLDSG 60
DB 1 MDVTIQHPMFARLTGPPYPSRLFPQFGEGLFXYDLPLFLSSTISPPYRQSLFRTVLDSG 60
QY 61 ISEVRSRDRKVFILDYKHSPEDLTAKVODDFVEIHGKHNERRDDHGYSISREFHRRYRL 120
DB 61 ISEVRSRDRKVFILDYKHSPEDLTAKVODDFVEIHGKHNERRDDHGYSISREFHRRYRL 120
QY 121 PSNVDSQALSCSLSDADGMLTFQCGKTIQTGLDATALAERALPVSREKPKPSAASS 173
DB 121 PSNVDSQALSCSLSDADGMLTFQCGKTIQTGLDATALAERALPVSREKPKPSAASS 173
DB 121 PSNVDSQALSCSLSDADGMLTFQCGKTIQTGLDATALAERALPVSREKPKPSAASS 173

RESULT 6
ID CRAA_EULFU STANDARD; PRT; 173 AA.
AC P02494;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha crystallin A chain.
GN CRYA.
OS Eulemur fulvus fulvus (Brown lemur).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Strepsirhini; Lemuridae; Eulemur.
OX NCBI_TaxID=40322;
RN [1]
RP PARTIAL SEQUENCE.
RA de Jong W.W., Zweers A., Goodman M.;
RT "Trends in the molecular evolution of alpha-crystallin.";
RL (In) Peeters H. (eds.);

```

RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
 CC Pergamon Press, Oxford (1980).
 CC -1- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 DR PIR: A02897; CYRPA.
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; Hsp20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin_1.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR ProDom: PD001193; Crystallin_N_1.
 DR PROSITE: PS01031; HSP20_1.
 KM Eye lens protein; Acetylation; Glycoprotein.
 FT MOD RES 1 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 SQ SEQUENCE 173 AA; 19819 MW; 288D3BD59B5BD37 CRC64;

Query Match 95.6%; Score 876; DB 1; Length 173;
 Best Local Similarity 94.8%; Pred. No. 4.7e-78;
 Matches 164; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHFWKRTIGPPYSLRFDQFEGGLFEYDLPELSTTSPYRQSLFRTVLDG 60
 DB 1 MDVTIQHFWKRTIGPPYSLRFDQFEGGLFEYDLPELSTTSPYRQSLFRTVLDG 60
 QY 61 ISEVSRDRKXVIFLDVGHFSPEDLTAVQDPEVHKGHNERODDHGYSREFHRRRL 120
 DB 61 ISEVSRDRKXVIFLDVGHFSPEDLTAVQDPEVHKGHNERODDHGYSREFHRRRL 120
 QY 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 QY 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173

RESULT 7
 CRAA_OCHPR STANDARD; PRT; 173 AA.
 AC P02492;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRYAA.
 OS Ocotona princeps (Southern American pika).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Ochotoniidae; Ochotona.
 OX NCBI_TaxID=9978;
 RN [1]
 RP PARTIAL SEQUENCE.
 RA de Jong W.W., Zweers A., Goodman M.;
 RT "Trends in the molecular evolution of alpha-crystallin.",
 RL (in) Peeters H. (eds.);
 RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
 RL Pergamon Press, Oxford (1980).
 CC -1- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 CC PIR: A02895; CYOIAA.
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; Hsp20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin_1.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR ProDom: PD001193; Crystallin_N_1.
 DR PROSITE: PS01031; HSP20_1.
 KM Eye lens protein; Acetylation; Glycoprotein.

FT MOD RES 1 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 SQ SEQUENCE 173 AA; 19780 MW; 77B8CAFB749A3A5C CRC64;

Query Match 95.3%; Score 873; DB 1; Length 173;
 Best Local Similarity 94.8%; Pred. No. 9.2e-78;
 Matches 164; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHFWKRTIGPPYSLRFDQFEGGLFEYDLPELSTTSPYRQSLFRTVLDG 60
 DB 1 MDVTIQHFWKRTIGPPYSLRFDQFEGGLFEYDLPELSTTSPYRQSLFRTVLDG 60
 QY 61 ISEVSRDRKXVIFLDVGHFSPEDLTAVQDPEVHKGHNERODDHGYSREFHRRRL 120
 DB 61 ISEVSRDRKXVIFLDVGHFSPEDLTAVQDPEVHKGHNERODDHGYSREFHRRRL 120
 QY 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVDSALSCSLSDGMLTFPCGPKIQGLDATHAERAIPIVSRREKPTSPASS 173

RESULT 8
 CRAA_PERPO STANDARD; PRT; 173 AA.
 AC P02495;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRYAA.
 OS Perodicticus potto edwardsi (Potto), and
 OS Galago crassicaudatus (Thick-tailed galago) (Orelemur crassicaudatus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Strepsirrhini; Loridae; Perodicticus.
 OX NCBI_TaxID=9473; 9463;
 RN [1]
 RP PARTIAL SEQUENCE.
 RC SPECIES=P.edwardsi, and G.crassicaudatus;
 RA de Jong W.W., Zweers A., Goodman M.;
 RT "Trends in the molecular evolution of alpha-crystallin.",
 RL (in) Peeters H. (eds.);
 RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
 RL Pergamon Press, Oxford (1980).
 RN [2]
 RP SEQUENCE OF 121-163 FROM N.A.
 RX MEDLINE=96139023; Pubmed=8587135;
 RA Jaworski C.V.;
 RT "A reassessment of mammalian alpha A-crystallin sequences using DNA
 RT sequencing: implications for anthropoid affinities of tarsier";
 RL J. Mol. Evol. 41:901-908(1995).
 CC -1- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 or send an email to license@ebi.ac.uk).

EMBL: U24068; AAA97570.1; -.
 PIR: A02898; CYPAA.
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; Hsp20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin_1.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR ProDom: PD001193; Crystallin_N_1.

DR PROSITE; PS01031; HSP20; 1.
 KW Eye lens protein; Acetylation; Glycoprotein.
 FT MOD RES 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 SQ SEQUENCE 173 AA; 19805 MW; 3318079BB5BD37 CRC64;

Query Match 95.3%; Score 873; DB 1; Length 173;
 Best Local Similarity 94.2%; Pred. No. 9.2e-78;
 Matches 163; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHWPFRKTLGPFYPSRLFDQFFGGLFEYDLPLFLSTTSPYRQSLFRTVLDG 60
 DB 1 MDVTIQHWPFRKTLGPFYPSRLFDQFFGGLFEYDLPLFLSTTSPYRQSLFRTVLDG 60
 QY 61 ISEVSRSDRDKVIFLDVHGFSPEDLTIVKQDPEVIEHGKHNERQDDHGYSRFEHRRYRL 120
 DB 61 VSEVRSDRDKVIFLDVHGFSPEDLTIVKQDPEVIEHGKHNERQDDHGYSRFEHRRYRL 120
 QY 121 PSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSPASS 173
 DB 121 PSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSPASS 173

RESULT 9

ID CRAA PROCA STANDARD; PRT; 173 AA.
 AC P02499;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 36, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRYAA.
 OS Procaeva capensis (Cape hyrax) (Rock dassie).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Hyacoridae; Procaviidae; Procavia.
 CX NCBI_TaxID=9813;
 RN [1]
 RN SEQUENCE.
 RA MEDLINE=77158093; PubMed=870070;
 RA de Jong W.W., Nuy-Terwindt E.C., Versteeg M.;
 RT "Primary structures of alpha-crystallin A chains of elephant, whale,
 RT hyrax and rhinoceros.";
 RL Biochim. Biophys. Acta 491:573-580(1977).
 CC -1- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 CC PIR: A02902; CYHXA.
 DR InterPro; IPR001436; Crystallin_alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin_1.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS; PRO0299; ACRYSTALLIN.
 DR ProDom; PD001193; Crystallin_N_1.
 DR PROSITE; PS01031; HSP20; 1.
 KW Eye lens protein; Acetylation; Glycoprotein.
 FT MOD RES 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT VARIANT 55 55 T -> A (TN 50% OF THE MOLECULES).
 SQ SEQUENCE 173 AA; 19822 MW; 96E0937224297228 CRC64;

Query Match 95.3%; Score 873; DB 1; Length 173;
 Best Local Similarity 93.6%; Pred. No. 9.2e-78;
 Matches 162; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHWPFRKTLGPFYPSRLFDQFFGGLFEYDLPLFLSTTSPYRQSLFRTVLDG 60
 DB 1 MDVTIQHWPFRKTLGPFYPSRLFDQFFGGLFEYDLPLFLSTTSPYRQSLFRTVLDG 60
 QY 61 ISEVSRSDRDKVIFLDVHGFSPEDLTIVKQDPEVIEHGKHNERQDDHGYSRFEHRRYRL 120
 DB 61 ISEVSRSDRDKVIFLDVHGFSPEDLTIVKQDPEVIEHGKHNERQDDHGYSRFEHRRYRL 120

DB 61 ISEVSRSDRDKVIFLDVHGFSPEDLTIVKQDPEVIEHGKHNERQDDHGYSRFEHRRYRL 120
 QY 121 PSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSPASS 173
 DB 121 PSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSPASS 173

RESULT 10

ID CRAA MOUSE STANDARD; PRT; 173 AA.
 AC P02490; P82532; Q61444;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alpha crystallin A chain, major component.
 GN CRYAA OR CRYA1.
 OS Mus musculus (Mouse).
 OS Rattus norvegicus (Rat).
 OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil),
 OS Mesocricetus auratus (Golden hamster), and
 OS Tupia glis (Tree shrew).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090, 10116, 10047, 10036, 9395;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=Rat; STRAIN=Sprague-Dawley;
 RA Bhat S.P., Nandy P., Srinivasan A., Cheng D., Sitey A.;
 RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RN SEQUENCE.
 RP SPECIES=Rat;
 RX MEDLINE=90105511; PubMed=2294971;
 RA Hendriks W., Weetink H., Voorter C.E.M., Sander J., Bloemendal H.,
 RA de Jong W.W.;
 RT "The alternative splicing product alpha Ains-crystallin is
 RT structurally equivalent to alpha A and alpha B subunits in the rat
 RT alpha-crystallin aggregate.";
 RL Biochim. Biophys. Acta 1037:58-65(1990).
 RN [3]
 RN PARTIAL SEQUENCE.
 RP SPECIES=Rat;
 RA de Jong W.W., van der Oudergraaf F.J., Versteeg M., Groenewoud G.,
 RA van Amelsvoort J.M., Bloemendal H.;
 RT "Primary structures of the alpha-crystallin A chains of seven
 RT mammalian species.";
 RL Eur. J. Biochem. 53:237-242(1975).
 RN [4]
 RN SEQUENCE OF 53-173 FROM N.A.
 RP SPECIES=Rat;
 RX MEDLINE=82081811; PubMed=6171772;
 RA Woormann R.J.M., van der Velden H.M.W., Dodemont H.J., Andreoli P.M.,
 RA Bloemendal H., Schoenmakers J.G.G.;
 RT "An unusually long non-coding region in rat lens alpha-crystallin
 RT messenger RNA.";
 RL Nucleic Acids Res. 9:4813-4822(1981).
 RN [5]
 RN SEQUENCE OF 90-172 FROM N.A.
 RP SPECIES=Rat;
 RX MEDLINE=93054670; PubMed=1429679;
 RA Srinivasan A.N., Nagineni C.N., Bhat S.P.;
 RT "Alpha A-crystallin is expressed in non-ocular tissues.";
 RL J. Biol. Chem. 267:23337-23341(1992).
 RN [6]
 RN SEQUENCE FROM N.A.
 RP SPECIES=Mouse;
 RX MEDLINE=83155647; PubMed=6187470;
 RA King C.R., Platiogorsky J.;
 RT "Alternative RNA splicing of the murine alpha A-crystallin gene:
 RT protein-coding information within an intron.";
 RL Cell 32:707-712(1983).
 RN [7]
 RN SEQUENCE OF 11-173 FROM N.A.

RC SPECIES=Mouse;
 RX MEDLINE=83119986; PubMed=7156978;
 RA King C.R., Shinozaki T., Patilgoreky J.;
 RT "Alpha A-crystallin messenger RNA of the mouse lens: more noncoding
 RT than coding sequences.";
 RL Science 215:985-987(1982).
 RN [8]
 RP PARTIAL SEQUENCE.
 RC SPECIES=M.unguiculatus, and M.auratus;
 RA de Jong W.W., Zweers A., Versteeg M., Nuy-Terwindt E.C.;
 RT "Trends in the molecular evolution of alpha-crystallin.";
 RL (In) Peeters H. (eds.);
 RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
 RL Pergamon Press, Oxford (1980).
 RN [9]
 RP SEQUENCE.
 RC SPECIES=T.gilis; TISSUE=Lens;
 RX MEDLINE=84208008; PubMed=6723655;
 RA de Jong W.W., Zweers A., Versteeg M., Nuy-Terwindt E.C.;
 RT "Primary structures of the alpha-crystallin A chains of twenty-eight
 RT mammalian species, chicken and frog.";
 RL Eur. J. Biochem. 141:131-140(1984).
 CC -I- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Major;
 CC IsoId=P02490-1; Sequence=Displayed;
 CC Name=Minor;
 CC IsoId=P02490-2; Sequence=Not described;
 CC -I- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U47922; AAA93367.1; -
 DR EMBL: V01219; AAA24530.1; -
 DR EMBL: M96949; AAA40644.1; -
 DR EMBL: M96950; AAA40645.1; -
 DR EMBL: J00376; AAA37471.1; -
 DR EMBL: V00730; AAA24108.2; -
 DR EMBL: J00375; AAA37469.1; -
 DR PIR: A02892; CYR1A.
 DR PIR: A02893; CYR1A.
 DR PIR: C94432; CYHAB.
 DR PIR: D94432; CYHAB.
 DR GLYCOSULEDB: P02490; -
 DR MGD: MGI:88515; Cryaa.
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; HSP20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin; 1.
 DR Pfam: PF00011; HSP20; 1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR ProDom: PD001193; Crystallin_N; 1.
 DR PROSITE: PS01031; HSP20; 1.
 DR Eye lens protein; Acetylation; Alternative splicing; Glycoprotein.
 FT MOD RES 1 ACETYLATION.
 FT CARBOHYD 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CONFLICT 124 124 V -> M (IN REF. 5).
 FT SEQUENCE 173 AA; 19792 MW; E146E3F488591F93 CRC64;
 Query Match 95.1%; Score 871; DB 1; Length 173;
 Best local similarity 94.8%; Pred. No. 1.4e-77;
 Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHWPFRKRTIGFPYPSRLPDQFEGEGFEYDLPFLSSTISPPYRQSLFRTVLDG 60
 DB 1 MDVTIQHWPFRKRTIGFPYPSRLPDQFEGEGFEYDLPFLSSTISPPYRQSLFRTVLDG 60
 QY 61 ISEVRSRDRKVFILVDVGHSPEDLTIVYQDDFVFIHKKNERODDGHYISREFFRRRL 120
 DB 61 ISEVRSRDRKVFILVDVGHSPEDLTIVYQDDFVFIHKKNERODDGHYISREFFRRRL 120
 QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGIDATHAERAIPIVSRREKPSAPSS 173
 DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGIDATHAERAIPIVSRREKPSAPSS 173
 RESULT 11
 CRAA_ARTUA
 ID CRAA_ARTUA STANDARD; PRT; 173 AA.
 AC P02482; 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRYAA.
 OS Artibeus jamaicensis (Jamaican fruit-eating bat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
 OC Stenodermatinae; Artibeus.
 OC NCBI_TaxID=9417;
 RN [1]
 RP PARTIAL SEQUENCE.
 RA de Jong W.W., Zweers A., Goodman M.;
 RT "Trends in the molecular evolution of alpha-crystallin.";
 RL (In) Peeters H. (eds.);
 RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,
 RL Pergamon Press, Oxford (1980).
 CC -I- FUNCTION: May contribute to the transparency and refractive index
 CC of the lens.
 CC -I- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 CC family.
 CC
 DR PIR: A02884; CYBTA.
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; HSP20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; crystallin; 1.
 DR Pfam: PF00011; HSP20; 1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR ProDom: PD001193; Crystallin_N; 1.
 DR PROSITE: PS01031; HSP20; 1.
 DR Eye lens protein; Acetylation; Glycoprotein.
 FT MOD RES 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT SEQUENCE 173 AA; 19807 MW; DB3A807BE01793FF CRC64;
 Query Match 94.8%; Score 869; DB 1; Length 173;
 Best local similarity 93.6%; Pred. No. 2.3e-77;
 Matches 162; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MDVTIQHWPFRKRTIGFPYPSRLPDQFEGEGFEYDLPFLSSTISPPYRQSLFRTVLDG 60
 DB 1 MDVTIQHWPFRKRTIGFPYPSRLPDQFEGEGFEYDLPFLSSTISPPYRQSLFRTVLDG 60
 QY 61 ISEVRSRDRKVFILVDVGHSPEDLTIVYQDDFVFIHKKNERODDGHYISREFFRRRL 120
 DB 61 ISEVRSRDRKVFILVDVGHSPEDLTIVYQDDFVFIHKKNERODDGHYISREFFRRRL 120
 QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGIDATHAERAIPIVSRREKPSAPSS 173
 DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGIDATHAERAIPIVSRREKPSAPSS 173
 RESULT 12
 CRAA_BOVIN
 ID CRAA_BOVIN STANDARD; PRT; 173 AA.

AC P02470; 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRVA OR CRVA1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 CX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88049675; PubMed=3675580;
 RA Hay R.E., Petrash J.M.;
 RT "Nucleotide sequence of a bovine lens alpha A-crystallin cDNA";
 RL Biochem. Biophys. Res. Commun. 148:31-37(1987).
 RN [2]
 RP SEQUENCE.
 RX MEDLINE=74075721; PubMed=4770792;
 RA van der Ouderaa F.J., de Jong W.W., Bloemendal H.;
 RT "The amino-acid sequence of the alphaA2 chain of bovine
 alpha-crystallin.";
 RL Eur. J. Biochem. 39:207-222(1973).
 RN [3]
 RP SEQUENCE, ACETYLATION, AND PHOSPHORYLATION.
 RX MEDLINE=91253722; PubMed=2042736;
 RA Smith J.B., Thevenon-Emeric G., Smith D.L., Green B.;
 RT "Elucidation of the primary structures of proteins by mass
 spectrometry.";
 RL Anal. Biochem. 193:118-124(1991).
 RN [4]
 RP CARBOHYDRATE-LINKAGE SITE SER-162.
 RX MEDLINE=92112709; PubMed=1730617;
 RA Rognemore E.P., Dell A., Morris H.R., Panico M., Reason A.J.;
 RA Savoy L.A., Wisow G.J., Zigler J.S. Jr., Earles B.J., Hart G.W.;
 RT "Vertebrate lens alpha-crystallins are modified by O-linked N-
 acetylglucosamine.";
 RL J. Biol. Chem. 267:555-563(1992).
 CC -!- FUNCTION: May contribute to the transparency and refractive index
 of the lens.
 CC -!- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M26142; AAA30471.1; -.
 DR PIR; A29656; CYBAA.
 DR GlycositeDB; P02470; -.
 DR InterPro; IPR001436; Crystallin alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; Crystallin; 1.
 DR Pfam; PF00011; HSP20; 1.
 DR PRINTS; PR00299; ACRYSTALLIN.
 DR ProDom; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 DR Eye lens protein; Acetylation; Phosphorylation; Glycoprotein.
 KW MOD_RES 122 122 ACETYLATION.
 FT MOD_RES 122 122 PHOSPHORYLATION.
 FT CARBOHYD 162 162 O-LINKED (GLCNAC).
 FT CARBOHYD 162 162 /FTID=CAR 000056.
 SQ SEQUENCE 173 AA; 19730 MW; 7796ED1B71864478 CRC64;

Query Match 94.9%; Score 869; DB 1; Length 173;
 Best Local Similarity 94.2%; Pred. No. 2.3e-77;

Matches 163; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MDVTIQHFWFKRTIGPPYSLRFDQFGEGLFEYDLPLFLSTTSPYRQSLFRVLDG 60
 DB 1 MDIAIQHFWFKRTIGPPYSLRFDQFGEGLFEYDLPLFLSTTSPYRQSLFRVLDG 60
 QY 61 ISEVRSDDKVFIFLDVGHFSPEDLTVAQDPEVHGHKNERODDGHYISREFFRRYL 120
 DB 61 ISEVRSDDKVFIFLDVGHFSPEDLTVAQDPEVHGHKNERODDGHYISREFFRRYL 120
 QY 121 PSNVDSALSCSLSDAGMLTFGPKIQTGDLTAERAIPVSRREKPTAPSS 173
 DB 121 PSNVDSALSCSLSDAGMLTFGPKIQTGDLTAERAIPVSRREKPTAPSS 173
 RESULT 13
 ID CRAA_TAPIN STANDARD; PRT; 173 AA.
 AC P02476;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha crystallin A chain.
 GN CRVA.
 OS Tapirus indicus (Asiatic tapir) (Malayan tapir).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.
 CX NCBI_TaxID=9802;
 RN [1]
 RP PARTIAL SEQUENCE.
 RA de Jong W.W., Zweers A., Goodman M.;
 RT "Trends in the molecular evolution of alpha-crystallin.";
 RL (In) Peeters H. (eds.);
 RL Proceedings of the biological fluids, Proc. 28th colloquium, pp.161-164,
 RL Pergamon Press, Oxford (1980).
 CC -!- FUNCTION: May contribute to the transparency and refractive index
 of the lens.
 CC -!- SIMILARITY: Belongs to the small heat shock protein (HSP20)
 family.
 CC -----
 DR PIR; A02878; CYTPAA.
 DR InterPro; IPR001436; Crystallin alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; Crystallin; 1.
 DR Pfam; PF00011; HSP20; 1.
 DR PRINTS; PR00299; ACRYSTALLIN.
 DR ProDom; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 DR Eye lens protein; Acetylation; Glycoprotein.
 KW MOD_RES 1 1 ACETYLATION (PROBABLE).
 FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CARBOHYD 162 162
 SQ SEQUENCE 173 AA; 19804 MW; 7629BD05806942DE CRC64;

Query Match 94.8%; Score 868; DB 1; Length 173;
 Best Local Similarity 93.6%; Pred. No. 2.8e-77;
 Matches 162; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHFWFKRTIGPPYSLRFDQFGEGLFEYDLPLFLSTTSPYRQSLFRVLDG 60
 DB 1 MDIAIQHFWFKRTIGPPYSLRFDQFGEGLFEYDLPLFLSTTSPYRQSLFRVLDG 60
 QY 61 ISEVRSDDKVFIFLDVGHFSPEDLTVAQDPEVHGHKNERODDGHYISREFFRRYL 120
 DB 61 ISEVRSDDKVFIFLDVGHFSPEDLTVAQDPEVHGHKNERODDGHYISREFFRRYL 120
 QY 121 PSNVDSALSCSLSDAGMLTFGPKIQTGDLTAERAIPVSRREKPTAPSS 173
 DB 121 PSNVDSALSCSLSDAGMLTFGPKIQTGDLTAERAIPVSRREKPTAPSS 173

RESULT 14
 CRAA_BALAC

```

ID  CRAA BALAC  STANDARD;  PRT;  173 AA.
AC  P02474;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Alpha crystallin A chain.
GN  CRYAA.
OS  Balaenoptera acutorostrata (Minke whale) (lesser rorqual).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC  Balaenopteridae; Balaenoptera.
OX  NCBI_TaxID=9767;
RN  [1]
RP  PARTIAL SEQUENCE.
RX  MEDLINE=77158093; PubMed=870070;
RA  de Jong W.W., Nuy-Terwindt E.C., Versteeg M.;
RT  "Primary structures of alpha-crystallin A chains of elephant, whale,
RT  hyrax and rhinoceros."
RL  Biochim. Biophys. Acta 491:573-580 (1977).
CC  -1- FUNCTION: May contribute to the transparency and refractive index
CC  of the lens.
CC  -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC  family.
CC  PIR: A02876; CYWHA.
DR  InterPro: IPR001436; Crystallin.alpha.
DR  InterPro: IPR003090; Crystallin_N.
DR  InterPro: IPR002068; Hsp20.
DR  InterPro: IPR008978; HSP20_chap.
DR  Pfam: PF00525; Crystallin_1.
DR  Pfam: PF00011; HSP20; 1.
DR  PRINTS: PR00299; ACRYSTALLIN.
DR  ProDom: PD001193; Crystallin_N; 1.
DR  PROSITE: PS01031; HSP20; 1.
KW  Eye lens protein; Acetylation; Glycoprotein.
FT  MOD RES 1 162 1 ACETYLATION (PROBABLE).
FT  CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ  SEQUENCE 173 AA; 19778 MW; DA9A2BD25B19590F CRC64;

Query Match 94.3%; Score 864; DB 1; Length 173;
Best local Similarity 93.1%; Pred. No. 6.9e-77;
Matches 161; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 MDVTIQHPWKRTIGPPYPSRLFDQFGEGIFEYDLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVAIQHPWKRALGPPYPSRLFDQFGEGIFEYDLPFLSSTISPYRQSLFRTVLDG 60

QY 61 ISEVSRDRKFEVFLDVKHSPEDLTVKVODDFVEIHGKNERODDGYISREFHRRYRL 120
DB 61 ISEVSRDRKFEVFLDVKHSPEDLTVKVOEDFVEIHGKNERODDGYISREFHRRYRL 120

QY 121 PSNVDSALSCSLSDGMLTFGCGPKIQGLDATHAERAIPVSRREKPTSA PSS 173
DB 121 PSNVDSALSCSLSDGMLTFGCGPKIPVSGMDAGHSERAIPVSRREKPTSA PSS 173

RESULT 15
CRAA GIRCA STANDARD; PRT; 173 AA.
AC P02471;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha crystallin A chain.
GN CRYAA.
OS Giraffa camelopardalis (Giraffe), and
OS Hippopotamus amphibius (Hippopotamus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Giraffoidea;
OC Giraffidae; Giraffa.
OX NCBI_TaxID=9894, 9833;
RN [1]
RP PARTIAL SEQUENCE.
RA de Jong W.W., Zweers A., Goodman M.;

```

```

RT "Trends in the molecular evolution of alpha-crystallin.";
RL (In) Peeters H. (eds.);
RL Proteins of the biological fluids, Proc. 28th colloquium, pp.161-164,
RL Pergamon Press, Oxford (1980).
CC -1- FUNCTION: May contribute to the transparency and refractive index
CC of the lens.
CC -1- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC family.
CC PIR: A02873; CYGPA.
DR PIR: A94432; CYGPA.
DR InterPro: IPR001436; Crystallin.alpha.
DR InterPro: IPR003090; Crystallin_N.
DR InterPro: IPR002068; Hsp20.
DR InterPro: IPR008978; HSP20_chap.
DR Pfam: PF00525; Crystallin_1.
DR Pfam: PF00011; HSP20; 1.
DR PRINTS: PR00299; ACRYSTALLIN.
DR ProDom: PD001193; Crystallin_N; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Eye lens protein; Acetylation; Glycoprotein.
FT MOD RES 1 162 1 ACETYLATION (PROBABLE).
FT CARBOHYD 162 162 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 173 AA; 19760 MW; DB312BCCAAF59CE CRC64;

Query Match 94.3%; Score 864; DB 1; Length 173;
Best local Similarity 93.6%; Pred. No. 6.9e-77;
Matches 162; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 MDVTIQHPWKRTIGPPYPSRLFDQFGEGIFEYDLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVAIQHPWKRALGPPYPSRLFDQFGEGIFEYDLPFLSSTISPYRQSLFRTVLDG 60

QY 61 ISEVSRDRKFEVFLDVKHSPEDLTVKVODDFVEIHGKNERODDGYISREFHRRYRL 120
DB 61 ISEVSRDRKFEVFLDVKHSPEDLTVKVOEDFVEIHGKNERODDGYISREFHRRYRL 120

QY 121 PSNVDSALSCSLSDGMLTFGCGPKIQGLDATHAERAIPVSRREKPTSA PSS 173
DB 121 PSNVDSALSCSLSDGMLTFGCGPKIPVSGVDAGHSERAIPVSRREKPTSA PSS 173

Search completed: September 27, 2004, 13:20:31
Job time : 12 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:17:58 ; Search time 40 Seconds

(without alignments)
1364.616 Million cell updates/sec

Title: US-10-657-740-1

Sequence: 1 MDVTIQHPWFKRTIGFPYPS.....HAERAIPIVSRREKFTAPSS 173

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP rhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP xvirus:*
16: SP bacteriopl:*
17: SP archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	759	82.9	173	13	Q8QFUS
2	736	80.3	171	13	Q8QFUS
3	694	75.8	173	13	Q8QFUS
4	653	71.3	173	13	Q8QFUS
5	488	53.3	170	11	Q99ND2
6	488	53.3	175	11	Q99PF3
7	480	52.4	176	11	Q80X03
8	458.5	50.1	168	13	Q9PUR2
9	441.5	48.2	172	13	Q9DEU9
10	441.5	48.2	172	13	Q804F0
11	423.5	46.2	208	13	Q93592
12	294.5	32.2	186	5	Q9GN07
13	293	32.0	56	4	Q13684
14	293	32.0	185	5	Q18634
15	290	31.7	190	5	Q95P25
16	289.5	31.6	191	4	Q96C20

17	277.5	30.3	205	4	Q96E17	Q96E17 homo sapien
18	254	27.7	145	5	Q8MLQ9	Q8MLQ9 dirosophila
19	250.5	27.3	133	5	Q9GT43	Q9GT43 anopheles g
20	249	27.2	121	5	Q9GSR6	Q9GSR6 bombyx mori
21	247	27.0	192	5	Q44112	Q44112 artemia san
22	237.5	25.9	182	11	Q99PR8	Q99PR8 mus musculu
23	221	24.1	43	6	Q28790	Q28790 pan troglod
24	219	23.9	122	5	Q9BMR0	Q9BMR0 bombyx mori
25	216	23.6	176	5	Q86G69	Q86G69 dermaceror
26	211.5	23.1	177	5	Q17268	Q17268 brugia paha
27	204	22.3	43	6	Q28453	Q28453 leontopithe
28	203	22.2	43	6	Q28213	Q28213 cebus albi
29	203	22.2	152	5	Q93141	Q93141 brugia mala
30	199	21.7	48	6	Q97731	Q97731 macropus ru
31	197	21.5	43	6	Q29142	Q29142 tarsius syr
32	197	21.5	48	13	Q9YH46	Q9YH46 opisthocomu
33	197	21.5	205	5	Q01718	Q01718 sarcophaga
34	194	21.2	48	13	Q9YH48	Q9YH48 lophura nyc
35	193.5	21.1	42	6	Q28312	Q28312 colobus que
36	193	21.1	48	13	Q9YH49	Q9YH49 turdus meru
37	192	21.0	48	6	Q97732	Q97732 didelphis m
38	192	21.0	48	13	Q9YH47	Q9YH47 sterna fusc
39	191.5	20.9	203	13	Q7SIX9	Q7SIX9 xenopus lae
40	191	20.9	43	6	Q29132	Q29132 tupata glis
41	191	20.9	43	6	Q28784	Q28784 propthecus
42	189	20.6	82	11	Q92212	Q92212 mus musculu
43	188	20.5	48	13	Q9YH50	Q9YH50 crotophaga
44	187.5	20.5	149	5	Q20660	Q20660 caenotnabi
45	187	20.4	43	6	Q28764	Q28764 pteropus hy

ALIGNMENTS

RESULT 1

ID	Q8QFUS	PRELIMINARY;	PRT;	173 AA.
AC	Q8QFUS;			
DT	01-JUN-2002 (TREMURel. 21, Created)			
DT	01-JUN-2002 (TREMURel. 21, Last sequence update)			
DT	01-OCT-2003 (TREMURel. 25, Last annotation update)			
DE	Newt alpha A-crystallin.			
OS	Cynops pyrrhogaster (Japanese common newt).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.			
OX	NCBI_TaxID=8330;			
RN	[1]			
RF	SEQUENCE FROM N.A.			
RA	Mizuno N., Agata K., Sawada K., Mochii M., Eguchi G.;			
RT	"Expression of crystallin genes in embryonic and regenerating newt lenses."			
RL	Dev. Growth Differ. 0:0-0(2002).			
DR	EMBL; D86299; BAB85811.1; -			
DR	InterPro; IPR001436; Crystallin_alpha.			
DR	InterPro; IPR003090; Crystallin_N.			
DR	InterPro; IPR002068; Hsp20.			
DR	InterPro; IPR008978; HSP20 chap.			
DR	Pfam; PF00525; Crystallin_1.			
DR	Pfam; PF00011; HSP20; 1.			
DR	PRINTS; PR00299; ACRYSTALLIN.			
DR	PRODom; PD001193; Crystallin_N; 1.			
DR	PROSITE; PS01031; HSP20; 1.			
DR	SEQUENCE 173 AA; 19951 MW; D7D8C8DDCC97B858 CRC64;			

Query Match 82.9%; Score 759; DB 13; Length 173;

Best Local Similarity 78.0%; Pred. No. 4.1e-69;

Matches 15; Conservative 22; Mismatches 16; Indels 0; Gaps 0;

QY	1	MDVTIQHPWFKRTIGFPYPSRLDPDFGGLFEDLLPFISSRTISYYROSLRVTYDVG 60
DB	1	MDVTIQHPWFKRTIGFPYPSRLDPDFGGLFEDLLPFISSRTISYYROSLRVTYDVG 60
QY	61	ISEVRSRDKKFIILDVKHFSPEDLTAKVQDDFVEIHGKNERQDDHGYISREFFHRRYL 120

Db 61 ISEVNSGRKFKQTLLDVKHFSPEDEUSVAILDDIYVEIHGKSHDRÖDDHGYREHERKRYRL 120

Qy 121 PSNVQDSALSCISLADGMILFCGPILQTLDAATHAERAIPIVSRREKPTSAES 173

Db 121 PASVDQSSITTCALSTDGMILFTSGTMMQSSLDSSHERRPIPIVSRREKPTSAQSS 173

RESULT 2	
Q9W6T3	
ID Q9W6T3	PRELIMINARY;
	PRT;
	171 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Alpha A crystalin.
 OS Xenopus laevis (African clawed frog).
 OC Amphibia, Batrachia, Chordata, Craniata, Vertebrata; Eumleiostroni;
 OC Anphibia, Batrachia, Anura; Mesobatrachia, Pipoidae, Pipidae;
 XN xenopodinae; xenopus.
 NCBI_TaxID=8355;

RP SEQUENCE FROM N.A.
RX MEDLINE=99251119; PubMed=10234811;
RA Mizuno N., Wochi M., Takahashi T.C., Eguchi G., Okada T.S.
RT "Lens regeneration in Xenopus is not a mere repeat of lens
development," with respect to crystallin gene expression.";
RL Differentiation 64:143-149(1999).

```
DR InterPro: IPR0001436, Crystallin_alpha.  
DR InterPro: IPR0003090, Crystallin_N.  
DR InterPro: IPR0020688, HSP20.  
DR InterPro: IPR008978; HSP20_chap.  
DR Pfam: PF00525; Crystallin_1.  
DR Pfam: PF00011; HSP20_1.  
DR PRINTS: PR00259; ACRYSTALLIN.  
DR ProDom: PD001193; Crystallin_N_1.  
DR PROSITE, PS01031; HSP20_1.  
DR PROSITE, PS01031; HSP20_1.  
DR SEQUENCE /I AA; 19789 MW; 96190D9FC5B7E712 CRC64
```

Query Match	80.3%	Score 736;	DB 13;	Length 171;
Best Local Similarity	77.1%;	Pred. No. 8.9e-67;		
Matches 131;	Conservative	26;	Mismatches 13;	Indels 0;
				Gaps 0;

QY
|||
|||
MDVTIHPWKRRLTGFVIFISKRLEDFGFGGLEBVLHLLPSSIISEFIRQDLEFVLVDSG 60

Db
1 MDITIIQHPWKRRLTGFVIFENRLLFDQVFGSGMDFELPEFMSSTISPPYKQNIISGYLDSG 60

QY
|||
|||
ISEVSRDRDFEVLFLVKKHSPEDLTIVKQDQFVFIHCKHNERQDHDGYSIREFHRRL 120

Db
61 ISEVSRDRDFVINTLVKKHSPEDLSYKHDDFVFIHCKHNERQDHDGYSIREFHRRL 120

QY
|||
|||
PSNVDSALSCSLADQMTLFCGEPKIQIGLDAITHAKRALPISREKPKISA 170

Db
121 PSNMDDNSVASCITLSADGILLTFGEPKIQSMDSHSDRITLPVSKKESGSS 170

RESULT 3	
Q8UUZ6	
ID Q8UUZ6	PRELIMINARY; PRT; 173 AA

```
DT 01-MAR-2002 (TREMBLrel). 20, Created)
DT 01-MAR-2002 (TREMBLrel). 20, last sequence update)
DT 01-OCT-2003 (TREMBLrel). 25, last annotation update)
DT Alpha A crystallin.
```

OS *Brachydanio rerio* (Zebrafish) (Danio rerio)
 OS *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;*
 OC *Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;*
 OC *Cyprinidae; Danio.*
 OX NCBI_TaxID=7955;
 OX [1]
 RN SEQUENCE FROM N.A.

PC TISSUE=Jens;
RA Runkle S., Hill J., Kantorow M., Horvitz J., Posner M.;
RT "Cloning and characterization of zebrafish (Danio rerio) alpha A-
RT crystallin.";
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.

DR ZFIN: ZDB-GENE-020508-1; cxyaa.
DR Interpro: IPR001436; Crystallin_alpha
DR Interpro: IPR003090; Crystallin_N.
DR Interpro: IPR002068; Hsp20.
DR Interpro: IPR008978; HSP20_chap.
DR Pfam: PRF0555; Crystallin_1.
DR Pfam: PRF0011; HSP20_1.
DR Pfam: PRD0299; ACRYSTALLIN.
DR Pfam: PD001193; Crystallin_N_1.
DR Prosite: PS01031; HSP20_1.
DR SSquence 173 AA; 15714 MW; 301D744

Query Match	75.8%;	Score 694;	DB 13;	Length 173;
Best Local Similarity	72.2%;	Pred. No. 1.7e-62;		
Matches 127; Conservative	24;	Mismatches 19;	Indels 6;	Gaps 3;

[illegible]

RESULT 4	
Q9DEV0	
ID Q9DEV0	PRELIMINARY; PRT; 173 AA

DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DB Alpha-A crystalin
DB *Clarias fuscus* (Whiteported *Clarias*)
OS *Clarias*; *Metaroi*; *Chorata*; *Cranata*; *Vertebrata*; *Euteleostomi*;
OC *Eucaryota*; *Metaroi*; *Chorata*; *Cranata*; *Vertebrata*; *Euteleostomi*;
OC *Neopterygii*; *Neopterygii*; *Teleostei*; *Ostariophysi*; *Siluriformes*;
OC *Clariidae*; *Clariidae*;
NCBI TaxID=33541;

RP SEQUENCE FROM N.A.
RA Chiu S.-H., Yu C.-M.;
RT "Alpha crystallin of catfish eye lenses: cDNA and genomic analysis of
RT alpha-A and alpha-B.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBr; AY007974; AAG2386c.1; 1.
DR InterPro; IPR001436; Crystallin alpha.
DR InterPro; IPR003090; Crystallin_N.
DR InterPro; IPR002068; Hsp20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin_1.
DR Pfam; PF00011; HSP20.1.
DR PRINTS; PRC0299; ACRYSTALLIN.
DR ProDom; PD001193; Crystallin_N.1.
DR PROSITE; PS01032; HSP20.1.
SQ SEQUENCE 173 AA; 19705 MW; 93AEEDAD122C9724 CRC64;

Query Match	71.3%;	Score 653;	DB 13;	Length 173;
Best Local Similarity	70.5%;	Pred. No. 2.5e-58;		
Matches 124; Conservative	23;	Mismatches 23;	Indels 6;	Gaps 3;

```

QY      1 MDVTIHPMFKRTLGPPYPSRLFDQFGEGLEFDLLPPLSSTISPYRKSLLFRYLD-- 58
      ||:||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MDIAIQPMFKRTLS--NPSRLDQFEGGLHDDLPTASTISPSYRSLFRSFLDSS 58

```

```

QY 59 -SGISEVSDRDKFYIFLDVKHFSPEDELTVKQDDFVEIHGKNERODDHGYISSEFRR 117
DB 59 NSGISEVSDRDKFYIFLDVKHFSPEDELTVKQDDFVEIHGKNERODDHGYISSEFRR 118
QY 118 YRLPSNVQOSALSCLSDAGMLTFGCPKIQGLDATHARAIIPVSREKPTSAAPS 173
DB 119 YRLPSNVQOALITLTSADGLSPGPK-TGSXKYGQDRTIPTRDDKPNPAAS 173

RESULT 5
Q99ND2 PRELIMINARY; PRT; 170 AA.
AC Q99ND2.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alphab-crystallin (Fragment).
GN CRYAB.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nannospalax.
OX NCBI_TaxID=30637;
RN [1]
RP SEQUENCE FROM N.A.
RA Smulders R.H., P.H., van Dijk M.M.A., Hoevenaars S., Lindner R.A.,
RA Carver J.A., de Jong W.W.;
RT "The evolutionary fate of mole rat alpha-A-crystallin: a redundant but
RT indispensable eye lens protein?";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ272441; CAC33095.1; -
DR InterPro; IPR001436; Crystallin_alpha.
DR InterPro; IPR003090; Crystallin_N.
DR InterPro; IPR002068; HSP20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin; 1.
DR Pfam; PF00011; HSP20; 1.
DR PRINTS; PR00299; ACRYSTALLIN.
DR PRODOM; PD001193; Crystallin_N; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 170
SQ SEQUENCE 170 AA; 19643 MW; 6B8F2183C1B66BB CRC64;

Query Match 53.3%; Score 488; DB 11; Length 170;
Best Local Similarity 55.2%; Pred. No. 1.5e-41;
Matches 95; Conservative 34; Mismatches 33; Indels 10; Gaps 5;

QY 1 MDVTIOHPWEKRTIGPPY-PSRLFDQFEGELFEYDLPFLISSTISPYRQ--SLFR--T 55
DB 1 MDIAIHHPWIRPPFPSPSRFLFDQFEGELLESDFLFS-TSTLSPPYLRPPSPFRAPS 59
QY 56 VLDGISEVSDRDKFYIFLDVKHFSPEDELTVKQDDFVEIHGKNERODDHGYISSEF 115
DB 60 WIDTGLSEMMEKDRFVNLDVKHFSPEELKVKVLGVIEVHGHERQDHGFSRFEH 119
QY 116 RRYRLPSNVQOSALSCLSDAGMLTFGCPKIQGLDATHARAIIPVSREKPT 167
DB 120 RKRYIPADVPLPLTITSLSDGVLTVMGPKRQ---ASGERTIPITREKPT 167

RESULT 6
Q9EPF3 PRELIMINARY; PRT; 175 AA.
AC Q9EPF3.
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alphab-B-crystallin.
GN ALPBA-B-CRYSTALLIN.
OS Spalax judaei.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;

```

```

OC Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Anza population; TISSUE=Heart, and Lens;
RX MEDLINE=21142392; PubMed=11245977;
RA Avioli A., Joel A., Nevo E.;
RT "The lens protein alpha-B-crystallin of the blind subterranean mole-
RT rat: high homology with sighted mammals.";
RL Gene 264:45-49(2001).
DR EMBL; AJ293658; CAC01692.1; -
DR InterPro; IPR001436; Crystallin_alpha.
DR InterPro; IPR003090; Crystallin_N.
DR InterPro; IPR002068; HSP20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin; 1.
DR Pfam; PF00011; HSP20; 1.
DR PRINTS; PR00299; ACRYSTALLIN.
DR PRODOM; PD001193; Crystallin_N; 1.
DR PROSITE; PS01031; HSP20; 1.
SQ SEQUENCE 175 AA; 20105 MW; 3BD653E0898B612F CRC64;

Query Match 53.3%; Score 488; DB 11; Length 175;
Best Local Similarity 54.5%; Pred. No. 1.5e-41;
Matches 97; Conservative 35; Mismatches 34; Indels 12; Gaps 6;

QY 1 MDVTIOHPWEKRTIGPPY-PSRLFDQFEGELFEYDLPFLISSTISPYRQ--SLFR--T 55
DB 1 MDIAIHHPWIRPPFPSPSRFLFDQFEGELLESDFLFS-TSTLSPPYLRPPSPFRAPS 59
QY 56 VLDGISEVSDRDKFYIFLDVKHFSPEDELTVKQDDFVEIHGKNERODDHGYISSEF 115
DB 60 WIDTGLSEMMEKDRFVNLDVKHFSPEELKVKVLGVIEVHGHERQDHGFSRFEH 119
QY 116 RRYRLPSNVQOSALSCLSDAGMLTFGCPKIQGLDATHARAIIPVSREKPT 171
DB 120 RKRYIPADVPLPLTITSLSDGVLTVMGPKRQ---ASGERTIPITREKPTAATAAP 173

RESULT 7
Q80X03 PRELIMINARY; PRT; 176 AA.
AC Q80X03.
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha B-crystallin (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9211524; PubMed=1765091;
RA Bhat S.P., Horwitz J., Srinivasan A., Ding L.;
RT "Alpha B-crystallin exists as an independent protein in the heart and
RT in the lens.";
RL Eur. J. Biochem. 202:775-781(1991).
DR EMBL; S77142; AAP3196.1; -
DR InterPro; IPR001436; Crystallin_alpha.
DR InterPro; IPR003090; Crystallin_N.
DR InterPro; IPR002068; HSP20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00525; crystallin; 1.
DR Pfam; PF00011; HSP20; 1.
DR PRINTS; PR00299; ACRYSTALLIN.
DR PRODOM; PD001193; Crystallin_N; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 1
SQ SEQUENCE 176 AA; 20168 MW; AF689379A6591914 CRC64;

Query Match 52.4%; Score 480; DB 11; Length 176;
Best Local Similarity 53.7%; Pred. No. 9.9e-41;

```

Matches 95; Conservative 36; Mismatches 34; Indels 12; Gaps 6;

QY 2 DVTIOHPFKRTLGPY-PSRLPDQFEGHFEYDLPLFSLSTISPYRQ--SLFR--TV 56
 Db 3 DIAHHWIRPPFPFSPSLPFGFHEHLLSDLS- TATSLSPFLRPPSLRASM 61

QY 57 LDGISEVSRDKEVIFLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSRFR 116
 Db 62 IDTGISEMKEKDRPVNLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSRFR 121

QY 117 RYRLPSNVQASLSCISADGMLTFCGPKIQGLDGTAEALPVSREKP--TSAP 171
 Db 122 KYRLPADVDPLTITSSLSGVLITNPRKQ---ASGPKRTITREKAVIRAP 174

RESULT 8
 Q9PUR2 PRELIMINARY; PRT; 168 AA.

AC Q9PUR2; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha b crystallin.
 GN CRYAB.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RX MEDLINE=20011287; PubMed=10542326;
 RA Posner M., Kantorow M., Horwitz J.;
 RT "Cloning, sequencing and differential expression of alpha-b-crystallin
 in the zebrafish, danio rerio(1)";
 RL Biochim. Biophys. Acta 1447:271-277 (1999).
 DR ZFIN; ZDB-GENE-991119-2; cryab.
 DR InterPro; IPR001436; Crystallin alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; crystallin; 1.
 DR PRINTS; PR00299; ACRYSTALLIN.
 DR PRODOM; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 SQ SEQUENCE 168 AA; 19377 MW; 7E79BC517BCE23 CRC64;

Query Match 50.1%; Score 458.5; DB 13; Length 168;
 Best Local Similarity 52.0%; Pred. No. 1.4e-38;
 Matches 91; Conservative 33; Mismatches 32; Indels 19; Gaps 5;

QY 1 MDVTIOHPFKRTLGP-PPSRLLPDQFEGHFEYDLPLFSLSTISPYRQSLFRTV-- 57
 Db 1 MEISIOHPYRRPLPFPFPYRIFDQYFGEHLSDSD--PF-----SFFTFMYRPPYMR 53

QY 58 -----DSGISEVSRDKEVIFLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSR 112
 Db 54 PFSWDSGSEKRODRKRVINLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSR 113

QY 113 EPHRRYRLPSNVQASLSCISADGMLTFCGPKIQGLDGTAEALPVSREKP 167
 Db 114 EFRKRYKIPAGVDPGALITSSLSGVLITNLRHQDI-----LERSIPITCEKP 164

RESULT 9
 Q9DEU9 PRELIMINARY; PRT; 172 AA.

AC Q9DEU9; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha-B crystallin.
 OS Clarias fuscus (Whitespotted clarias).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
 OC Clariidae; Clarias.
 OC NCBI_TaxID=33541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chou S.-H., Yu C.-M., Chao Y.-K.;
 RT "Alpha crystallin of catfish eye lenses: cDNA and genomic analysis of
 alpha-A and alpha-B";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY007973; AAC23867.1; -
 DR InterPro; IPR001436; Crystallin alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; crystallin; 1.
 DR PRINTS; PR00299; ACRYSTALLIN.
 DR PRODOM; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 SQ SEQUENCE 172 AA; 19831 MW; 74DC93030396AB64 CRC64;

Query Match 48.2%; Score 441.5; DB 13; Length 172;
 Best Local Similarity 49.1%; Pred. No. 7.9e-37;
 Matches 86; Conservative 40; Mismatches 40; Indels 9; Gaps 5;

QY 1 MDVTIOHPFKRTLGP-PPSRLLPDQFEGHFEYDLPLFSLSTISPYRQSLFRTV--TVL 57
 Db 1 MDIAIOHPFRSRPFSPSLPFGFHEHLLSDLS- TATSLSPFLRPPSLRASM 58

QY 58 DSGISEVSRDKEVIFLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSRFR 117
 Db 59 ESGISEMKEKDRPVNLDVYHSPEDLTVKQDDPVEIHGKNERODDHGYSRFR 118

QY 118 YRLPSNVQASLSCISADGMLTFCGPKIQGLDGTAEALPVSREKPSAPS 172
 Db 119 YRVSGVDPLTITSSLSGVLITNPRKPS--DA--PERSTITREKAVIRAP 169

RESULT 10
 Q804E0 PRELIMINARY; PRT; 172 AA.

AC Q804E0; 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha-B crystallin.
 OS Clarias batrachus (walking catfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
 OC Clariidae; Clarias.
 OC NCBI_TaxID=59899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chou S.-H., Yu C.-M., Chao Y.-K.;
 RT "Alpha crystallin of catfish eye lenses: cDNA and genomic analysis of
 alpha-A and alpha-B";
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY184812; AAC24775.1; -
 DR InterPro; IPR001436; Crystallin alpha.
 DR InterPro; IPR003090; Crystallin_N.
 DR InterPro; IPR002068; HSP20.
 DR InterPro; IPR008978; HSP20_chap.
 DR Pfam; PF00525; crystallin; 1.
 DR PRINTS; PR00299; ACRYSTALLIN.
 DR PRODOM; PD001193; Crystallin_N; 1.
 DR PROSITE; PS01031; HSP20; 1.
 SQ SEQUENCE 172 AA; 19831 MW; 74DC93030396AB64 CRC64;

Query Match 48.2%; Score 441.5; DB 13; Length 172;
 Best Local Similarity 49.1%; Pred. No. 7.9e-37;
 Matches 86; Conservative 40; Mismatches 40; Indels 9; Gaps 5;

QY 1 MDVTIQHPWFKRTL-GFPYPSRLFPQFEGEFLFYDLPLPISSTISPYRQSLFR--TVL 57
 DB 1 MDIAIQHPWFKRSFQSFPSRIFPDQHFHSESEVALAPHSYVCP--RSQFPMPSWV 58
 QY 58 DSGISEVRSDRDKFVIFLDVHFSPELITVKVQDDFVEIHGKNERODHGYISREFRR 117
 DB 59 ESGISEMKEKDRFTINDVGHFTPEELGVKSGDIYEVHAKHEDRODDHGFVSSEFHRX 118
 QY 118 YRLSNVQGSALSCSLSDGMLTFGCPKIQGLDATHARALPVSREKPTSAQS 172
 DB 119 YRBSGVPTSTITSLSSDGVLTTPARKPS--DA--PERSITTRDKSKVSGSG 169

RESULT 11

Q93592 PRELIMINARY; PRT; 208 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha-A-crystallin.
 OS Astyanax fasciatus (Blind cave fish) (Astyanax mexicanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Characiformes;
 OC Characidae; Astyanax.
 CX NCB1_TaxID=7994;
 RX MEDLINE=98398454; PubMed=9729440;
 RA Behrens M., Wilkens H., Schmale H.;
 RT "Cloning of the alpha-crystallin genes of a blind cave form and the
 RT epigen form of Astyanax fasciatus: a comparative analysis of
 RT structure, expression and evolutionary conservation.";
 RL Gene 216:319-326 (1998).
 DR EMBL: Y11300; CAA72158.1;
 DR InterPro: IPR001436; Crystallin alpha.
 DR InterPro: IPR003090; Crystallin_N.
 DR InterPro: IPR002068; HSP20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00525; Crystallin; 1.
 DR Pfam: PF00011; HSP20; 1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR PRODOM: PD001193; Crystallin_N; 1.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 208 AA; 22568 MW; D32AE17C07244DB3 CRC64;

Query Match 46.2%; Score 423.5; DB 13; Length 208;
 Best Local Similarity 72.9%; Pred. No. 6.8e-35;
 Matches 78; Conservative 13; Mismatches 11; Indels 5; Gaps 2;

QY 1 MDVTIQHPWFKRTLGPYPSPRLPDQFEGEFLFYDLPLPISSTISPYRQSLFTVLD-- 58
 DB 1 MDIAIQHPWFKRRLG--YPSRLFPQFEGEFLFYDLPLPISSTISPYRSLFRNFDSS 58
 QY 59 -SGISEVRSDRDKFVIFLDVHFSPELITVKVQDDFVEIHGKNERQ 104
 DB 59 NSGISEVRSDRDKFVIFLDVHFSPEELNVKVAEDVYEQKIGERQ 105

RESULT 12

Q9GN07 PRELIMINARY; PRT; 186 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Heat shock protein hsp20.8A.
 HSP20.8A.

OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditryaia; Bombycoidea;
 OC Bombycidae; Bombyx.
 CX NCB1_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Li B., Xia Q.-Y., Fujii H., Banno Y.;
 RT "Isolation, identification and particular expression of small heat-
 RT shock genes in silkworm, Bombyx mori.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF15319; AAC30946.1;
 DR EMBL: AF15317; AAG30944.1;
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR002068; HSP20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00011; HSP20; 1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 186 AA; 20804 MW; F8956B919BF978CF CRC64;

Query Match 33.2%; Score 294.5; DB 5; Length 186;
 Best Local Similarity 41.8%; Pred. No. 7.7e-22;
 Matches 64; Conservative 28; Mismatches 48; Indels 13; Gaps 5;

QY 18 YPSRLFPQFEGEFLFYDL-----PLSSTISPYR--QSLFRTVLDGISEVRSDRK 70
 DB 19 WPSRLVQDGLALTPMDMLAAVACPLS---EDYFPWRRLAASNDLG--SLKADKX 74
 QY 71 FVIFLDVHFSPELITVKVQDDFVEIHGKNERODHGYISREFRRRLPSNVQSL 130
 DB 75 FQVNLVDQHFSPESISVKTADGYIVEGKHEKKDEHYSRQVRRYALBGAAPTE 134
 QY 131 CSLSADGMLTFGCPKIQGLDATHARALPYSR 163
 DB 135 SRLSSDGVLTTPARKVP--DAVGERKVP 165

RESULT 13

Q13684 PRELIMINARY; PRT; 56 AA.

DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha-A-crystallin (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Theria; Primates; Catarrhini; Homnidae; Homo.
 CX NCB1_TaxID=9606;
 RX MEDLINE=95137069; PubMed=7835394;
 RA Caspers G.J., Pennings J., De Jong W.W.;
 RT "A partial cDNA sequence corrects the human alpha A-crystallin primary
 RT structure.";
 RL Exp. Eye Res. 59:125-126 (1994).
 DR EMBL: L25781; AAC17570.1;
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR002068; HSP20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00011; HSP20; 1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR PROSITE: PS01031; HSP20; 1.
 FT NON_TER 1
 FT TER 56
 SQ SEQUENCE 56 AA; 6121 MW; 4898A44D349EA0F7 CRC64;

Query Match 32.0%; Score 293; DB 4; Length 56;
 Best Local Similarity 98.2%; Pred. No. 2.4e-22;
 Matches 55; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 105 DDHGYISREFRRRLPSNVQSLSCSLSDGMLTFGCPKIQGLDATHARALP 160

Db 1 DDHGYSREFFRRYRLPSNVQSLSCSLSDGMLTFCGPYKIQGLDATHAEALP 56

RESULT 14

018634 PRELIMINARY; PRT; 185 AA.
 ID 018634
 AC 018634;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Alpha-crystallin cognate protein 25.
 GN ACP25.
 OS Plodia interpunctella (Indianmeal moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditypsia; Pyraloidea;
 CC Pyralidae; Phycitinae; Plodia.
 CX NCBI_TaxID=58624;
 (1)
 RN SEQUENCE FROM N.A.
 RX MEDLINE=98318807; PubMed=9654738;
 RA Shirk P.D., Broza R., Hemphill M., Perera O.P.;
 RT "alpha-Crystallin protein cognates in eggs of the moth, Plodia
 interpunctella: possible chaparones for the follicular epithelium yolk
 protein.";
 RT Insect Biochem. Mol. Biol. 28:151-161 (1998).
 RL EMBL: U94328; AAC36146.1; -;
 DR InterPro: IPR001436; Crystallin_alpha.
 DR InterPro: IPR002068; Hsp20.
 DR InterPro: IPR008978; Hsp20_chap.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 185 AA; 20639 MW; AA492D095535C44C CRC64;

Query Match 32.0%; Score 293; DB 5; Length 185;

Best Local Similarity 41.2%; Pred.No.1.le-21;
 Matches 68; Conservative 25; Mismatches 48; Indels 24; Gaps 7;

Db 7 HPWFRTLGPFYPSRLFDQFEGGLFEYDLT-----PFLSTISPYR--QSLFRTVLD 59
 18 HSW-----PSRLIDNFGIALTPDEM LTA VACP LLS---TDYKRPWRQLAAARDI 65
 QY 60 GISEVSRDRKXVIFLDVKGHFSPEDLTVKQDPEVIEHGENERQDDHGYSREFFRRY 119
 Db 66 G-SNIKADRKQIMNDVGHFSPEISVKTADGFVVEGKHEKDEHGYSISRCQFVRRYA 124
 QY 120 LPSNVQSLSCSLSDGMLTFCGP-KIQGLDATHAEALPVS 163
 Db 125 LPEGAASETVESRLSSGGVLTITAPLKP--DAVGERKVPIDQ 166

RESULT 15

095P25 PRELIMINARY; PRT; 190 AA.

ID 095P25
 AC 095P25;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Heat shock protein HSP27-like protein.
 OS Cliona intestinalis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 CC Phlebobranchia; Clonidae; Clona.
 CX NCBI_TaxID=7719;
 (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Body wall muscle;
 RX MEDLINE=21100321; PubMed=11159910;
 RA Vandenberghe A.E., Meedel T.H., Hastings K.B.;
 RT "mRNA 5'-leader trans-splicing in the chordates.";
 RL Genes Dev. 15:294-303 (2001).
 DR EMBL: AF237691; AAL27684.1; -;
 DR InterPro: IPR001436; Crystallin_alpha.

DR InterPro: IPR002068; Hsp20.
 DR InterPro: IPR008978; HSP20_chap.
 DR Pfam: PF00011; HSP20_1.
 DR PRINTS: PR00299; ACRYSTALLIN.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 190 AA; 21416 MW; 131826BF7B591550 CRC64;

Query Match 31.7%; Score 290; DB 5; Length 190;

Best Local Similarity 36.7%; Pred.No.2.3e-21;
 Matches 66; Conservative 26; Mismatches 44; Indels 44; Gaps 5;

QY 10 EKRTLGPRY---PRLIDQFEGGLF--EYDLPLSLSTISPYRQSLR----- 54
 Db 17 YRSMDTFYDMPTSSLFPQDFGSAPMSADFLA-----KQRFDPMTARGRP 65
 QY 55 -----TVLDSGISEVSRDRKXVIFLDVKGHFSPEDLTVKQDPEVIEHGH 100
 Db 66 TAGVPIVPAPMTVRQLSGMSQVTTDENKFKVTLDVGHFTPEITVKTVDGALFVHGKH 125
 QY 101 NERQDDHGYSREFFRRYRLPSNVQSLSCSLSDGMLTFCGP---KIQGLDATHAE 156
 Db 126 HEKEDHGVSARDFTKRYTIPPNVDPLTVTSSLSPDGILTVAPIRALQAPTSLSVQHLE 185

Search completed: September 27, 2004, 13:21:18

Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:19:19 ; Search time 32 Seconds

(without alignments)
279.103 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVYIQHMFRTIGPFYPS.....HAERAIYVREKPTAPSS 173

Scoring table: BLOSUM62

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	299	32.6	193	2 US-08-900-407-3	Sequence 3, Appli
2	289.5	31.6	459	4 US-09-513-783A-170	Sequence 170, App
3	288	31.4	199	2 US-08-900-407-4	Sequence 4, Appli
4	283	30.9	232	4 US-09-553-498-6	Sequence 6, Appli
5	283	30.9	232	4 US-09-618-869-6	Sequence 6, Appli
6	197	21.5	60	4 US-09-621-976-7032	Sequence 7032, Ap
7	171.5	18.7	196	2 US-08-900-407-1	Sequence 1, Appli
8	96	10.5	139	4 US-09-107-532A-4324	Sequence 4324, Ap
9	92	10.0	154	4 US-09-252-991A-28607	Sequence 28607, A
10	92	10.0	154	4 US-09-489-039A-10634	Sequence 10634, A
11	76.5	8.4	191	4 US-09-489-039A-10634	Sequence 10634, A
12	76.5	8.3	156	3 US-08-459-953A-10	Sequence 10, Appl
13	76	8.3	586	4 US-09-393-212-10	Sequence 10, Appl
14	76	8.3	741	1 US-08-277-331A-4	Sequence 4, Appli
15	76	8.3	741	2 US-08-473-750-7	Sequence 7, Appli
16	76	8.3	741	2 US-08-477-326-7	Sequence 7, Appli
17	75	8.2	119	4 US-09-543-681A-7735	Sequence 7735, Ap
18	73.5	8.0	167	4 US-09-835-909A-2	Sequence 2, Appli
19	73	8.0	210	4 US-09-252-991A-23899	Sequence 23899, A
20	73	8.0	915	3 US-08-480-917-2	Sequence 2, Appli
21	73	8.0	915	3 US-09-138-736-2	Sequence 2, Appli
22	72	8.0	915	4 US-08-988-242-2	Sequence 2, Appli
23	72.5	7.9	389	1 US-08-650-275-3	Sequence 3, Appli
24	72.5	7.9	389	1 US-09-181-318-3	Sequence 3, Appli
25	72	7.7	553	4 US-09-465-559-14	Sequence 14, Appl
26	70.5	7.7	243	4 US-09-198-452A-284	Sequence 284, App
27	70.5	7.7	534	3 US-09-124-541-1	Sequence 1, Appli

28	70.5	7.7	534	4 US-09-663-326-1	Sequence 1, Appli
29	70	7.6	983	4 US-09-252-991A-21667	Sequence 21667, A
30	69.5	7.6	157	4 US-09-107-532A-4800	Sequence 4800, Ap
31	69.5	7.6	327	4 US-09-489-039A-8522	Sequence 8522, Ap
32	69.5	7.6	2074	4 US-09-491-356C-9	Sequence 9, Appli
33	69	7.5	159	4 US-09-732-210-68	Sequence 68, Appl
34	69	7.5	455	4 US-09-252-991A-18265	Sequence 18265, A
35	69	7.5	792	4 US-09-543-681A-6617	Sequence 6617, Ap
36	68.5	7.5	2409	6 5180808-2	Patent No. 5180808
37	68	7.4	147	4 US-09-107-532A-4808	Sequence 4808, Ap
38	68	7.4	277	4 US-09-252-991A-26048	Sequence 26048, A
39	68	7.4	989	4 US-09-079-030-218	Sequence 218, App
40	68	7.4	1838	3 US-09-120-663-2	Sequence 2, Appli
41	68	7.4	1838	4 US-09-431-614-8	Sequence 8, Appli
42	67.5	7.4	272	4 US-09-107-532A-5541	Sequence 5541, Ap
43	67.5	7.4	906	4 US-09-717-364A-15	Sequence 15, Appl
44	67	7.3	288	4 US-09-252-991A-27243	Sequence 27243, A
45	67	7.3	368	4 US-09-624-390-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-900-407-3
Sequence 3, Application US/08900407
Patent No. 5962262
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,407
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0351 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 63522
US-08-900-407-3
Query Match 32.6%; Score 299, DB 2; Length 193;
Best Local Similarity 42.7%; Pred. No. 6,9e-28;

Matches 73; Conservative 24; Mismatches 46; Indels 28; Gaps 5;

```

QY 1 FYXSRFLDDEFG-----EGFYXDLPLPLSSTL---SEYYSQSLFRT 55
Db 23 YHOSRFLRFDOSFGPMHPLPDMWYKPPSSGAMPGYFR--LPLPSALLPAGSYYGA---LS 77

QY 56 VLOSGLSEVSDRDKRVFLDVAKHSPEDLTVKVDDEFLHGKNEQODHGYISREFH 115
Db 78 ELSSGSISETQSDSKMKVLDVNHHPPELTVYKTDNVEIGKHEKQODEHGFLSRCT 13

QY 116 RRRRLPSNVDSALSCSLADGMLTFQCGPKIQTLGDATEARALPVSSEK 166
Db 138 RKTLPFGVATATVRSLSLSDGMLTVEALPLKPAQS--SETLITVYVEAK 186

```

RESULT 2

```

Sequence 170, Application US/09513783A
Patent No. 6416959
GENERAL INFORMATION:
APPLICANT: Giuliano, Kenneth A.
APPLICANT: Kapur, Ravi
TITLE OF INVENTION: A System for Cell Based Screening
FILE REFERENCE: 97-022-L1
CURRENT APPLICATION NUMBER: US/09/513,783A
CURRENT FILING DATE: 2000-02-25
NUMBER OF SEQ ID NOS: 180
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 170
LENGTH: 459
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP22
US-09-513-783A-170

```

Query Match	31.6%;	Score 289.5;	DB 4;	Length 459;
Best Local Similarity	37.9%;	Pred. No. 3.4e-26;		
Matches	77;	Conservative	24;	Mismatches 51;
				Indels 51;
				Gaps 7

```

QY 10 FKRTLGG-----FPP-SRLTDPQFPGEL.FEYDLDPFL-----40
Db 262 FSLIRGSMWDFPDWPHSHRLFPOAG-----LPRUPBMSOMLGSSMPGYAPLP 3133
QY 41 -----SSTISPYRROSLFETVLDSGISSEVRSDRDKFYFLDPKIFSPEDLTIVXQDFV 94
Db 314 PPAIESPAVAAPKXSRALSNQ-LSSQVSHIRHTADRWMRSLDVNHNFAVDDELTVTKDGvV 372
QY 95 EIHGKNERODHGYYISREFHRRYRLPSNVDSALSCISLADQMLTFGCEKIQGLDATH 154
Db 373 EITCKHEERODEHGYYISRCFTRKYTLPPVCPDPTOVSSLSPEGLTIVAPMKLATQSN- 431
QY 155 AERALPYRSREK-----PTSAPS 172
Db 432 -EITIVTFESRAOLGGPEPAKS 453

```

RESULT 3

US-08-900-407-4
Sequence 4, Application US/08900407
Patent No. 5962262
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto

STATE: CA
COUNTRY: USA

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: Diskette
3 COMPUTER: IBM Compatible
4 OPERATING SYSTEM: DOS
5 SOFTWARE: FASTSEO for Windows Version 2.0
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/900,407
8 FILING DATE: Filed Herewith
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER:
11

```

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME: Billings, Lucy J.
3  REGISTRATION NUMBER: 36,749
4  REFERENCE/DOCKET NUMBER: PF-0351 US
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE: 415-855-0555
7  TELEFAX: 415-845-4166
8  INFORMATION FOR SEO ID NO: 4:
9  SEQUENCE CHARACTERISTICS:
10 LENGTH: 199 amino acids
11 TYPE: amino acid
12 STRANDEDNESS: single
13 TOPOLOGY: linear
14 IMMEDIATE SOURCE:
15 LIBRARY: GenBank
16 CLONE: 662841
17
18 US-08-900-407-4

```

Query Match	31.4%;	Score 288;	DB 2;	Length 199;
Best Local Similarity	38.5%;	Pred. No. 1.5e-26;		
Matches	74;	Conservative	23;	Mismatches 49;
				Indels 46;
				Gaps 6;

```

QY 10 EKRLTG-----FYF-SRLDFQFGBGLFYEDLLPFL-----40
Db 8 FSLRLGSPWMDPRDDYFHSRLPFDQAFG-----LPLRLPEWSOMLGSSSMGYYRPLP 59
QY 41 -----SSTISPYKYQSLFRFTVLDSGISHEVRSRDRLKFVILDYKHSPEDLTVKQDDFV 94
Db 60 PAATESPAVAAPAYSRALSRQ-LSSGVSEIRHTADRMWRSLVYNHFAPEDLFTKDKGVV 118
QY 95 EIHGHKHNRRDDHGYSIRSFHRRYKLPNSVDQSALSCLSSADQMLTFCCPKIQTGLDATH 154
Db 119 EITGNGHRSQEHGYSIRSCFTRKTYLPSPGVDPDVQSSLSPEGTLVLEAPMPKLATQSN- 177
QY 155 AERATVSGREEK 166
Db 178 -EITLPVTFESR 188

```

RESULT 4

```

US-09-553-498-6
; Sequence 6, Application US//09553498
; Patent No. 6109861
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorothee
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: Process for the production of naturally folded and secreted proteins
; FILE REFERENCE: Case 20379
; CURRENT APPLICATION NUMBER: US//09/553,498
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: EP99107412.1
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 6
; LENGTH: 232
; TYPE: PRT
; ORGANISM: E. coli
;

```

US-09-553-498-6

Query Match 30.9%; Score 283; DB 4; Length 232;
Best Local Similarity 40.1%; Pred. No. 7.7e-26;
Matches 75; Conservative 21; Mismatches 43; Indels 48; Gaps 7;

QY 16 PF---YR--SRLPDQFG-----EGLEFYDLPLPLS 41
DB 41 PFRDWPFAHSRLFPDAGFVPRLPDEWSQWFSAGMGVYRPLPAATAG-----PAAY 93
QY 42 STISPYRQSLFRTVLDSGISVRSDBRDKFVFLDVKHSPEBDLVKVQDDFVEIHGKN 101
DB 94 TLAPAFSRAINRQ--LSSGVSEIRQTADRWVSLDVNHFAPBELTVKKEGVETIGKHE 152
QY 102 ERDDHGYISREHRRRLPSNVDSALSCSLADGMLTFGC--PKIQGLDATHAERAI 159
DB 153 ERQDEHGYISRCFRKRYTLPPGVDPFLVSSLSPEGLTVEAPLPKAVT-----QSAEITTI 208
QY 160 PVSREBK 166
DB 209 PVTFEAR 215

RESULT 5
US-09-618-869-6

; Sequence 6, Application US/09618869
; Patent No. 6435279
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorthree
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwartz, Elisabeth
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND
; TITLE OF INVENTION: SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR
; FILE REFERENCE: 20381
; CURRENT APPLICATION NUMBER: US/09/618, 869
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: EP9911481.5
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-618-869-6

Query Match 30.9%; Score 283; DB 4; Length 232;
Best Local Similarity 40.1%; Pred. No. 7.7e-26;

Matches 75; Conservative 21; Mismatches 43; Indels 48; Gaps 7;

QY 16 PF---YR--SRLPDQFG-----EGLEFYDLPLPLS 41
DB 41 PFRDWPFAHSRLFPDAGFVPRLPDEWSQWFSAGMGVYRPLPAATAG-----PAAY 93
QY 42 STISPYRQSLFRTVLDSGISVRSDBRDKFVFLDVKHSPEBDLVKVQDDFVEIHGKN 101
DB 94 TLAPAFSRAINRQ--LSSGVSEIRQTADRWVSLDVNHFAPBELTVKKEGVETIGKHE 152
QY 102 ERDDHGYISREHRRRLPSNVDSALSCSLADGMLTFGC--PKIQGLDATHAERAI 159
DB 153 ERQDEHGYISRCFRKRYTLPPGVDPFLVSSLSPEGLTVEAPLPKAVT-----QSAEITTI 208
QY 160 PVSREBK 166
DB 209 PVTFEAR 215

RESULT 6
US-09-621-976-7032
; Sequence 7032, Application US/09621976
; Patent No. 6639063

GENERAL INFORMATION:

APPLICANT: Dumas Mline Edwards, J.B.

APPLICANT: Jobert, S.

APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: ESTS and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2

CURRENT APPLICATION NUMBER: US/09/621,976

NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent.pm

SEQ ID NO 7032

LENGTH: 60

TYPE: PRT

ORGANISM: Homo sapiens

US-09-621-976-7032

Query Match 21.5%; Score 197; DB 4; Length 60;
Best Local Similarity 61.1%; Pred. No. 2.4e-16;
Matches 33; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

QY 94 VEIHGKNERQDDHGYISREHRRRLPSNVDSALSCSLADGMLTFGC PKIQ 147
DB 3 IEVHGKNERQDEHGFISREHRRRLPSNVDSALSCSLADGMLTFGC PKIQ 56

RESULT 7

US-08-900-407-1

Sequence 1, Application US/08900407

Patent No. 5962262

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer

APPLICANT: Lal, Preeti

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Puryi

TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE

TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/900,407

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0351 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-845-4166

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 196 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: ?????

CLONE: 1362715

US-08-900-407-1

Query Match 18.7%; Score 171.5; DB 2; Length 186;
 Best Local Similarity 31.2%; Pred. No. 1.6e-12;
 Matches 45; Conservative 24; Mismatches 58; Indels 17; Gaps 2;

20 SRLFDQFEGSLFEYIDL-----LPLSTSTSPYRQSLFRI-----VLDSGIS 62
 Db 28 SRLLDGFGMDPFDDDLTASWPMALPRLSSAMPGLTRSGVPRGPATATAREGVPARGRT 87
 QY 63 EVRSDRKKEVIFLIVKHPSPEDLTIVKQVDDPVEIHGKHNERODDHGYISREHRRYRLPS 122
 Db 88 PPPFGEPMKVCVAVHSEKPEELMWKTKDGYEVSGHNEKQEGEGLVSKNFTKKIDLP 147
 QY 123 NVDSALSCSISADGMLTFGCPKI 146
 Db 148 EVDPTVFASISPEGLLITIEAPQV 171

RESULT 8
 US-09-107-532A-4324
 ; Sequence 4324, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESS: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107, 532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085, 598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-8277
 TELEFAX: (781)893-5007
 INFORMATION FOR SEQ ID NO: 4324:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 139 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (8) LOCATION 1..139
 SEQUENCE DESCRIPTION: SEQ ID NO: 4324:
 US-09-107-532A-4324

Query Match 10.5%; Score 96; DB 4; Length 139;
 Best Local Similarity 25.0%; Pred. No. 0.0012;
 Matches 31; Conservative 25; Mismatches 46; Indels 22; Gaps 6;

35 DILPLSTSTSPYRQSLFRIYLDGSI-----EVRSDDRKEVIFLIVKHPSPEDLTIVK 89

Db 11 DMPDFDFVSPAPNDFL-----GVSSYPKVDLVNEKEKYLITADMPGDKEDTIVEX 63

QY 90 QDDFVEI---HGKNERODDHGYISREH-----RRYRLPSNVDSALSCSISADGMLTF 141
 Db 64 SDNLTITSAHNEHSTEDKEKGNVYRKERHVSYSYRSFYL NVDEKITGTFR NGVILKL 121

QY 142 CGPK 145
 Db 122 VLPK 125

RESULT 9
 US-09-252-991A-28807
 ; Sequence 28807, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252, 991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 28807
 LENGTH: 154
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-28807

Query Match 10.0%; Score 92; DB 4; Length 154;
 Best Local Similarity 24.6%; Pred. No. 0.0041;
 Matches 34; Conservative 22; Mismatches 50; Indels 32; Gaps 5;

QY 8 PWEKRTGAPFYPSRLFDQFEGSLFEYIDLPLPLSTSTSPYRQSLFRIYLDGSIENVSD 67
 Db 14 PLFRHSVG-----PREF--NDLFESALRNEAGSTVPPY-----NVEKHG 50
 QY 68 RDKVYIFLDVAKHSPEDLTIVKQVDDPVEIHGKHNERODH-----GYISREHRRYRLP 121
 Db 51 DDERIVIAANGFOEBDDLVYRGVLTVSGCKRKSIDNTYTHGIAQAFKLSFRLA 110
 QY 122 SNVDQALSCSISADGML 139
 Db 111 DHIEVKASL---ANGTL 125

RESULT 10
 US-09-489-10634
 ; Sequence 10634, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 APPLICANT: Gary Breton et. al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 2709.2004001
 CURRENT APPLICATION NUMBER: US/09/489, 039A
 CURRENT FILING DATE: 2000-01-27
 PRIOR APPLICATION NUMBER: US 60/117,747
 PRIOR FILING DATE: 1999-01-29
 NUMBER OF SEQ ID NOS: 14342
 SEQ ID NO 10634
 LENGTH: 191
 TYPE: PRT
 ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-10634

Query Match 10.0%; Score 92; DB 4; Length 191;
 Best Local Similarity 19.9%; Pred. No. 0.0056;

[illegible]

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
? MEDIUM TYPE: storage
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: IBM P.C. DOS 5.0
? SOFTWARE: Microsoft Word
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/459,953A
? FILING DATE: June 2, 1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/029,494
? FILING DATE: March 19, 1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Wardburg, Richard J.
? REGISTRATION NUMBER: 32,327
? REFERENCE/DOCKET NUMBER: 209/267
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (213) 489-1600
? TELEFAX: (213) 955-0440
? TELEX: 67-3510
? INFORMATION FOR SEQ ID NO: 10 :
? SEQUENCE CHARACTERISTICS:
? LENGTH: 586 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-459-953A-10

Query Match      8.3%; Score 76; DB 3; Length 586;
Best Local Similarity 27.8%; Pred. No.2.5;
Matches 44; Conservative 18; Mismatches 56; Indels 40; Gaps 10;

QY      33 EYDLPLPSTISIPYRQSLFRVLDGISEVRSRDRDKFVFLDPK----HFSPEDLTVK 88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      138 ETDLARLEQGTALAEHAKLFMYQLLGLKTIHSAN--VIHRDLKPANITISTEDLVK 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      89 VQD---DVEYIHKNNERODDHGYISR---EFHRRYRL--PSN---VDGALSGSL 133
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      195 IGDGLRIAYDQHSR-----GYLSEGLVTWKYRSPRLILSPNNYTKALIDMAAGCIL 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      134 S--ADGMLTFCG---PKIQTGLDATTAEKAIPIYSRE 165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      249 AEMLTGRMLFPAHAHELEOMQITL-----TTPVIREE 280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-393-212-10
; Sequence 10, Application US/09393212
; Patent No. 6579972
; GENERAL INFORMATION:
; APPLICANT: Lechner, Cornelia
;            Moller, Niels P.H.
;            Ulrich, Axel
; TITLE OF INVENTION: EXTRACELLULAR SIGNAL-RELATED
;                    KINASE, SEQUENCES, AND
;                    METHODS OF PRODUCTION
;                    AND USE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 613 West Fifth Street
;           Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; STORAGE
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Microsoft Word

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/393,212
FILING DATE: 09-Sep-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/459,953A
FILING DATE: June 2, 1995
APPLICATION NUMBER: 08/029,494
FILING DATE: March 19, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/267
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 10 :
SEQUENCE CHARACTERISTICS:
LENGTH: 586 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-393-212-10

Query Match 8.3%; Score 76; DB 4; Length 586;
Best Local Similarity 27.8%; Pred. No. 2.5;
Matches 44; Conservative 18; Mismatches 56; Indels 40; Gaps 10;

QY 33 EYDLPFISSTISPYRQSLFRVYDSDGISEVRSRDKPFVFLDVK---HSPEDLVTK 88
DB 138 ETDRLARLEQCTLAEBNAKLFMYQLRLKXIHSMN--VLRDLKPNIFISTEDLVK 194
QY 89 VQD---DFVEIHGKHNRQDDHGYSR---EPRRYRL---PSN---VDQSALSGSL 133
DB 195 IGDGLARIVQOHSHK-----GYLSGLVTKWTRSPRLLSPPNYTKALIDMMAAGCIL 248
QY 134 S-ADGMLTFCG---PKIQGLDNTAHERAIPVSRRE 165
DB 249 AEMLTGRMLFAGAHLEQMOJILE-----TIPVIREE 280

RESULT 14
US-08-277-231A-4
Sequence 4, Application US/08277231A
Patent No. 5643725
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP Pili Operon of No. 5643725typable
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS: Haemophilus Influenzae
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,231A
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.

REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: ACC94-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-277-231A-4

Query Match 8.3%; Score 76; DB 1; Length 741;
Best Local Similarity 24.2%; Pred. No. 3.5;
Matches 38; Conservative 20; Mismatches 57; Indels 42; Gaps 5;

QY 20 SRLPDPFGGLPFYDLPFLP---SSTISPYRQSLFRVYDSDGISEVRSRDKPFVFL 75
DB 275 SRTDENVLQGVLYGLNHLTNSILYTRHYAGLFGFGLNTPIGAFSADATWSHAF 334
QY 76 DVKHPSPEDLVTKVQDDFVEIHGKHNRQDDHG-----YISRPH-----RR 117
DB 335 PLKHSKNGYS-----LHGSYSINFNESGTNITLAAYRVSSRDPFYTLSDTIGLNR 385
QY 118 YR-----LPSNVDSALSCSISADGMLTFCG 143
DB 386 FRQSPGAVLPETYPKQFVLSQSLGNNGLYLSG 422

RESULT 15
US-08-473-750-7
Sequence 7, Application US/08473750
Patent No. 5834187
Patent No. 5834187 5786143
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5834187
TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5834187 5786143
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS: Haemophilus Influenzae
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,750
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,321
FILING DATE: 19-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 amino acids

; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-473-750-7

Query Match 8.3%; Score 76; DB 2; Length 741;
 Best Local Similarity 24.2%; Pred. No. 3.5;
 Matches 38; Conservative 20; Mismatches 57; Indels 42; Gaps 5;

 QY 20 SRLPDQFFGGGLFEYDILPLF-----SSTISPPYRQSLFRVYLDGISEVRSDDRDKFYFL 75
 Db 275 SRTEDEVVLQGVLYGLTNHTLNSSLLYRHRARGLFGRLNTPIGAFSDATWSHAFF 334
 QY 76 DVKHFSPEDLTVKVQDDFVEIHGKNERQDDG-----YISREFH-----RR 117
 Db 335 PLKHTVSKNGYS-----LHGSYSINFNESGTNITLAAYRYSSRDFTLSDTIGLNRT 385
 QY 118 YR-----LPSNVDSALSCSLSDGMLTFCC 143
 Db 386 FROPSGAYLPETIYRPKNQFVLSQSLGNWGNLYLSG 422

Search completed: September 27, 2004, 13:22:17
 Job time : 33 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:20:24 ; Search time 48 Seconds

(without alignments)
1158.951 Million cell updates/sec

US-10-657-740-1

Title:

Perfect score: 916
Sequence: 1 MDVTIQHMPFKRTIGPPYPS.....HAERALPVSREKPTSPASS 173

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched:

1349238 seqs, 321558718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	173	16	US-10-657-740-1
2	871	95.1	173	14	US-10-316-253-113
3	869	94.9	173	16	US-10-657-740-10
4	849.5	92.7	196	14	US-10-316-253-111
5	849.5	92.7	196	16	US-10-657-740-12
6	683	74.6	177	13	US-10-105-427-2
7	639	69.8	123	16	US-10-657-740-3
8	531	58.0	105	16	US-10-657-740-18
9	493	54.5	175	16	US-10-657-740-11
10	489	53.4	175	16	US-10-408-765A-405
11	485	52.9	175	12	US-10-205-331-104
12	340	37.0	160	14	US-10-226-956-297
13	340	37.0	160	16	US-10-468-091-19
14	339	37.1	160	14	US-10-226-956-299
15	338	36.9	63	14	US-10-029-386-27955

16	338	36.9	160	14	US-10-226-956-298	Sequence 238, App
17	336	36.7	162	16	US-10-468-091-20	Sequence 20, Appl
18	289.5	31.6	205	12	US-09-969-034-4479	Sequence 4479, App
19	289.5	31.6	205	14	US-10-153-668-284	Sequence 284, App
20	288.5	31.6	205	15	US-10-116-755-148	Sequence 148, App
21	289.5	31.6	205	16	US-10-408-765A-435	Sequence 435, App
22	289.5	31.6	459	14	US-10-100-957A-170	Sequence 170, App
23	288.5	31.5	205	14	US-10-316-253-109	Sequence 109, App
24	288	31.4	199	16	US-10-408-765A-41	Sequence 41, Appl
25	288	31.4	471	10	US-09-935-642-4	Sequence 4, Appl
26	284.5	31.1	102	12	US-09-990-47-729	Sequence 4, Appl
27	283	30.9	209	16	US-10-657-740-9	Sequence 29, Appl
28	277.5	30.3	205	16	US-10-408-765A-2305	Sequence 2305, App
29	266	29.0	136	16	US-10-468-091-18	Sequence 18, Appl
30	188.5	20.6	143	15	US-10-369-493-6423	Sequence 6423, App
31	188.5	20.6	143	15	US-10-369-493-6424	Sequence 6424, App
32	154.5	16.9	145	16	US-10-657-740-7	Sequence 7, Appl
33	150	16.4	145	15	US-10-369-493-6422	Sequence 6422, App
34	150	16.4	145	15	US-10-369-493-6425	Sequence 6425, App
35	147	16.0	186	16	US-10-408-765A-2017	Sequence 2017, App
36	115	12.6	159	15	US-10-104-047-3656	Sequence 3656, App
37	110.5	12.1	245	15	US-10-104-047-2942	Sequence 2942, App
38	107.5	11.7	153	14	US-10-298-638-12	Sequence 12, Appl
39	106.5	11.6	207	16	US-10-767-701-44859	Sequence 44859, A
40	105	11.5	160	12	US-10-424-599-14696	Sequence 14696, A
41	103	11.2	195	12	US-10-424-599-14697	Sequence 14697, A
42	101.5	11.1	208	15	US-10-310-154-692	Sequence 692, App
43	101	11.0	161	12	US-10-424-599-161876	Sequence 161876, App
44	100	10.9	231	12	US-10-424-599-176081	Sequence 176081, A
45	99	10.8	153	12	US-10-424-599-161149	Sequence 161149, A

ALIGNMENTS

RESULT 1
US-10-657-740-1
Sequence 1, Application US/10657740
Publication No. US20040157289A1
GENERAL INFORMATION:
APPLICANT: Salerno, John C.
APPLICANT: Hanna, Michael
APPLICANT: Koretz, Jane F.
APPLICANT: Creone, Donna
APPLICANT: Smith, Susan E.
TITLE OR INVENTION: PROTEIN EXPRESSION SYSTEM
FILE REFERENCE: 01794100406US1
CURRENT APPLICATION NUMBER: US/10/657,740
CURRENT FILING DATE: 2003-09-08
PRIOR APPLICATION NUMBER: US 60/408,680
PRIOR FILING DATE: 2002-09-06
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 173
TYPE: PRT
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: GenBank / P02489
DATABASE ENTRY DATE: 1986-07-21
RELEVANT RESIDUES: (1)..(173)
US-10-657-740-1

Query Match 100.0%; Score 916; DB 16; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.4e-96;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVTIQHMPFKRTIGPPYPSRLPDQFFGEGLFYDLPFLSTSTSPYRQSLPRTYLDG 60
DB 1 MDVTIQHMPFKRTIGPPYPSRLPDQFFGEGLFYDLPFLSTSTSPYRQSLPRTYLDG 60
QY 61 ISEVRSRDKFVFIETLVKHFSPEDLTGVYQDDFVEIHKGNRQDDHGYSREHRRYRL 120

Db 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173
Db 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173

RESULT 2

US-10-316-253-113
; Sequence 113, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-316-253-113

Query Match 95.1%; Score 871; DB 14; Length 173;
Best Local Similarity 94.8%; Pred. No. 2e-91;
Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHWFKRTIGPEYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
Db 1 MDVTIQHWFKRALGPPYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
QY 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
Db 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173
Db 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173

RESULT 3

US-10-657-740-10
; Sequence 10, Application US/10657740
; Publication No. US20040157289A1
; GENERAL INFORMATION:
; APPLICANT: Salerno, John C.
; APPLICANT: Hanna, Michael
; APPLICANT: Koretz, Jane F.
; APPLICANT: Smith, Susan E.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; FILE REFERENCE: 01794100H406US1
; CURRENT APPLICATION NUMBER: US/10/657,740
; CURRENT FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: US 60/408,680
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-657-740-10

Query Match 94.9%; Score 869; DB 16; Length 173;

Best Local Similarity 94.2%; Pred. No. 3.4e-91;
Matches 163; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHWFKRTIGPEYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
Db 1 MDVTIQHWFKRTIGPEYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
QY 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
Db 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173
Db 121 PSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHERAIPVSRREKPTSAFSS 173

RESULT 4

US-10-316-253-111
; Sequence 111, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 111
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-316-253-111

Query Match 92.7%; Score 849.5; DB 14; Length 196;
Best Local Similarity 83.7%; Pred. No. 7.1e-89;
Matches 164; Conservative 5; Mismatches 4; Indels 23; Gaps 1;

QY 1 MDVTIQHWFKRTIGPEYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
Db 1 MDVTIQHWFKRALGPPYPSRLFDQFEGGLFEYDLPEFLSTISPYRQSLFRTVLDSG 60
QY 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
Db 61 ISEVSRDRKRVILDVYKHSPEDLTVKVDVFEIHKHNERODDHGYISREFFRRYRL 120
QY 98 GKHNERODDHGYISREFFRRYRLPSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHER 157
Db 121 GKHNERODDHGYISREFFRRYRLPSNVDSALSCSLSDGMLTFGCPKIQTGDLTAHER 180
QY 158 AIPVSRREKPTSAFSS 173
Db 181 AIPVSRREKPTSAFSS 196

RESULT 5

US-10-657-740-12
; Sequence 12, Application US/10657740
; Publication No. US20040157289A1
; GENERAL INFORMATION:
; APPLICANT: Salerno, John C.
; APPLICANT: Hanna, Michael
; APPLICANT: Koretz, Jane F.
; APPLICANT: Smith, Susan E.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; FILE REFERENCE: 01794100H406US1

```

; CURRENT APPLICATION NUMBER: US/10/657,740
; CURRENT FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: US 60/408,680
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-657-740-12
```

```

Query Match          92.7%; Score 849.5; DB 16; Length 196;
Best Local Similarity 83.7%; Pred. No. 7.1e-89;
Matches 164; Conservative 5; Mismatches 4; Indels 23; Gaps 1;
```

```

QY      1 MDVTIOHPWFKRTGPFYPSRLFDQFGEGLFEYDLPPLSGTISPYRQ--SLPR--T 55
Db      1 MDVTIOHPWFKRALGPFYPSRLFDQFGEGLFEYDLPPLSGTISPYRQSLFRTVLD 60
QY      61 ISB-----VSRDKFVIFLDVYKHFSPEDLTVKVOODFVEIH 97
Db      61 ISELMTMWFVMDQPHAGNPKNPVKRSRDKFVIFLDVYKHFSPEDLTVKVLDFEIH 120
QY      98 GKNHERODDHGYISREHRRYRLPSNVDSALSCSLADGMLTFCGPKIQTGLDATAER 157
Db      121 GKNHERODDHGYISREHRRYRLPSNVDSALSCSLADGMLTFCGPKVSGGLDAGHSER 180
QY      158 AIPVSRREKPTAPSS 173
Db      181 AIPVSRREKPSAPSS 196
```

RESULT 6

```

US-10-105-427-2
; Sequence 2, Application US/10105427
; Publication No. US20020177192A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: Chimeric protein alpha BNAC crystallin with extraordinarily high
; FILE REFERENCE: US- 649
; CURRENT APPLICATION NUMBER: US/10/105,427
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of chimeric protein alpha BNAC
US-10-105-427-2
```

```

Query Match          74.6%; Score 683; DB 13; Length 177;
Best Local Similarity 75.3%; Pred. No. 7.7e-70;
Matches 134; Conservative 16; Mismatches 22; Indels 6; Gaps 4;
```

```

QY      1 MDVTIOHPWFKRTGPFY-PSRLFDQFGEGLFEYDLPPLSGTISPYRQ--SLPR--T 55
Db      1 MDVAIHHPWTKRPPFPSPSRLLFDQFGEHLBESDLP--TSTLSPEFYLRPPSFLRAPS 59
QY      56 VLDSGISSEVSRDCKFVIFLDVYKHFSPEDLTVKVOODFVEIHGKNHERODDHGYISREH 115
Db      60 WFTGTGSEMKLEKDRFVSVDVKNHFSPEDLTVKVOODFVEIHGKNHERODDHGYISREH 119
QY      116 RRRRLPSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTAPSS 173
Db      120 RRRRLPSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTAPSS 177
```

RESULT 7

```
US-10-657-740-3
```

```

; Sequence 3, Application US/10657740
; Publication No. US20040157289A1
; GENERAL INFORMATION:
; APPLICANT: Salerno, John C.
; APPLICANT: Hanna, Michael
; APPLICANT: Koretz, Jane F.
; APPLICANT: Crone, Donna
; APPLICANT: Smith, Susan E.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; FILE REFERENCE: 01794100H406US1
; CURRENT APPLICATION NUMBER: US/10/657,740
; CURRENT FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: US 60/408,680
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-657-740-3
```

```

Query Match          69.8%; Score 639; DB 16; Length 123;
Best Local Similarity 100.0%; Pred. No. 5.2e-65;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      51 SLPRTVDSGISSEVSRDCKFVIFLDVYKHFSPEDLTVKVOODFVEIHGKNHERODDHGYI 110
Db      1 SLPRTVDSGISSEVSRDCKFVIFLDVYKHFSPEDLTVKVOODFVEIHGKNHERODDHGYI 60
QY      111 SREHRRYRLPSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSA 170
Db      61 SREHRRYRLPSNVDSALSCSLADGMLTFCGPKIQTGLDATAERAIPVSRREKPTSA 120
QY      171 PSS 173
Db      121 PSS 123
```

RESULT 8

```

US-10-657-740-18
; Sequence 18, Application US/10657740
; Publication No. US20040157289A1
; GENERAL INFORMATION:
; APPLICANT: Salerno, John C.
; APPLICANT: Hanna, Michael
; APPLICANT: Koretz, Jane F.
; APPLICANT: Crone, Donna
; APPLICANT: Smith, Susan E.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; FILE REFERENCE: 01794100H406US1
; CURRENT APPLICATION NUMBER: US/10/657,740
; CURRENT FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: US 60/408,680
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-657-740-18
```

```

Query Match          58.0%; Score 531; DB 16; Length 105;
Best Local Similarity 96.2%; Pred. No. 1e-52;
Matches 101; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```

QY      45 SPYRQSLPRTVDSGISSEVSRDCKFVIFLDVYKHFSPEDLTVKVOODFVEIHGKNHERO 104
Db      1 SPYRQSLPRTVDSGISSEVSRDCKFVIFLDVYKHFSPEDLTVKVOODFVEIHGKNHERO 60
QY      105 DDHGYISREHRRYRLPSNVDSALSCSLADGMLTFCGPKIQTG 149
```



```

; SOFTWARE: Patentin version 3.1
; SEQ ID NO 297
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-10-226-956-297

Query Match          37.1%; Score 340; DB 14; Length 160;
Best Local Similarity 40.1%; Pred. No. 1.5e-30;
Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

QY 3 VTIOHPWFKRTLTGPF-----YPSRLFDQFEGGLFEYDLPFLSSTISPYRQSLFRTVLD 58
D 5 VPQPSWLRRAADLPGLSAPGRLPDQRFEGGLLEALALCPTTLAPY---LRAPVA 61

QY 59 SGISEVSRDRDKFVIFLDVGHFSPEDLTIVKQDDFVEIHGKNRQDDHGYSREFFRRY 118
D 62 LPVAQVPTDPGHFSLVDVGHFSPDELTAVKVGSHVHARHERPDEHGFVAREFFRRY 121

QY 119 RLPSNVDSALSCSLSDGMLTFCGPKIQGLDATHAERAI PVSRREKPTSA 170
D 122 RLPGVDPAAVTALSPEGVLSI-----QAAPASQAAPPRA 158

RESULT 13
US-10-468-091-19
; Sequence 19, Application US/10468091
; Publication No. US20040157329A1
; GENERAL INFORMATION:
; APPLICANT: ADP Pharmaceutical Pty Limited
; TITLE OF INVENTION: The University of Sydney
; FILE REFERENCE: 500311
; CURRENT APPLICATION NUMBER: US/10/468,091
; PRIOR FILING DATE: 2003-08-13
; PRIOR APPLICATION NUMBER: AU PR3116
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 19
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-468-091-19

Query Match          37.1%; Score 340; DB 16; Length 160;
Best Local Similarity 40.1%; Pred. No. 1.5e-30;
Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

QY 3 VTIOHPWFKRTLTGPF-----YPSRLFDQFEGGLFEYDLPFLSSTISPYRQSLFRTVLD 58
D 5 VPQPSWLRRAADLPGLSAPGRLPDQRFEGGLLEALALCPTTLAPY---LRAPVA 61

QY 59 SGISEVSRDRDKFVIFLDVGHFSPEDLTIVKQDDFVEIHGKNRQDDHGYSREFFRRY 118
D 62 LPVAQVPTDPGHFSLVDVGHFSPDELTAVKVGSHVHARHERPDEHGFVAREFFRRY 121

QY 119 RLPSNVDSALSCSLSDGMLTFCGPKIQGLDATHAERAI PVSRREKPTSA 170
D 122 RLPGVDPAAVTALSPEGVLSI-----QAAPASQAAPPRA 158

RESULT 14
US-10-226-956-299
; Sequence 299, Application US/10226956
; Publication No. US20030060399A1
; GENERAL INFORMATION:
; APPLICANT: Biophy, Colleen
; APPLICANT: Komalavilas, Padmini
; APPLICANT: Panitch, Alyssa
; APPLICANT: Joshi, Lokesh

```

```

; APPLICANT: Seal, Brandon L.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR SMOOTH MUSCLE THERAPIES
; FILE REFERENCE: ASU-1061-US
; CURRENT APPLICATION NUMBER: US/10/226,956
; PRIOR FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 60/314,535
; NUMBER OF SEQ ID NOS: 320
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 299
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-10-226-956-299

Query Match          37.0%; Score 339; DB 14; Length 160;
Best Local Similarity 40.1%; Pred. No. 2e-30;
Matches 69; Conservative 31; Mismatches 50; Indels 22; Gaps 3;

QY 3 VTIOHPWFKRTLTGPF-----YPSRLFDQFEGGLFEYDLPFLSSTISPYRQSLFRTVLD 58
D 5 VPQPSWLRRAADLPGLSAPGRLPDQRFEGGLLEALALCPTTLAPY---LRAPVA 61

QY 59 SGISEVSRDRDKFVIFLDVGHFSPEDLTIVKQDDFVEIHGKNRQDDHGYSREFFRRY 118
D 62 LPVAQVPTDPGHFSLVDVGHFSPDELTAVKVGSHVHARHERPDEHGFVAREFFRRY 121

QY 119 RLPSNVDSALSCSLSDGMLTFCGPKIQGLDATHAERAI PVSRREKPTSA 170
D 122 RLPGVDPAAVTALSPEGVLSI-----QAAPASQAAPPRA 158

RESULT 15
US-10-029-386-27955
; Sequence 27955, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27955
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21 69.0
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: P02493, EVALUUE 2.00e-32
US-10-029-386-27955

Query Match          36.9%; Score 338; DB 14; Length 63;
Best Local Similarity 100.0%; Pred. No. 7e-31;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVTIOHPWFKRTLTGPFYPSRLFDQFEGGLFEYDLPFLSSTISPYRQSLFRTVLD 60
D 1 MDVTIOHPWFKRTLTGPFYPSRLFDQFEGGLFEYDLPFLSSTISPYRQSLFRTVLD 60

```

QY	61	ISE	63
Db	61	ISE	63

Search completed: September 27, 2004, 13:23:11
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2004, 13:18:24 ; Search time 16 Seconds
(without alignments)
1040.071 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916
Sequence: 1 MDVTIQHPWFRRTIGPPYPS.....HAERALPVSRKEKPTASS 173

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR.78:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	173	1	CYHDA
2	884	96.5	173	1	CYELAA
3	883	96.4	173	1	CYRBA
4	879.5	96.0	172	1	CYMOA
5	878	95.9	173	1	CYCPA
6	878	95.9	173	1	CYCPA
7	876	95.6	173	1	CYELAA
8	873	95.3	173	1	CYOLPA
9	873	95.3	173	1	CYOLPA
10	873	95.3	173	1	CYGCMA
11	873	95.3	173	1	CYHXA
12	871	95.1	173	1	CYHXA
13	871	95.1	173	1	CYHXA
14	871	95.1	173	1	CYHXA
15	871	95.1	173	1	CYHXA
16	869	94.9	173	1	CYBOA
17	869	94.9	173	1	CYBOA
18	868	94.8	173	1	CYTPA
19	864	94.3	173	1	CYHXA
20	864	94.3	173	1	CYHXA
21	864	94.3	173	1	CYHXA
22	862	94.1	173	1	CYHXA
23	860	93.9	173	1	CYHXA
24	859	93.8	173	1	CYHXA
25	857	93.6	173	1	CYHXA
26	856	93.4	173	1	CYHXA
27	856	93.4	173	1	CYHXA
28	856	93.4	173	1	CYHXA
29	854	93.2	173	1	CYHXA

30	850	92.8	173	1	CYHXA	alpha-crystallin c
31	850	92.8	195	1	CYHXA	alpha-crystallin c
32	850	92.8	195	1	CYHXA	alpha-crystallin c
33	849.5	92.7	196	2	S07530	alpha-crystallin c
34	844	92.1	173	1	CYHXA	alpha-crystallin c
35	844	92.1	173	1	CYHXA	alpha-crystallin c
36	843	92.0	173	1	CYHXA	alpha-crystallin c
37	839	91.6	173	1	CYHXA	alpha-crystallin c
38	831	90.7	173	1	CYHXA	alpha-crystallin c
39	829.5	90.6	170	1	CYHXA	alpha-crystallin c
40	816.5	89.1	170	1	CYHXA	alpha-crystallin c
41	816	89.1	173	1	CYHXA	alpha-crystallin c
42	806.5	88.0	196	2	A28332	alpha-crystallin c
43	803.5	87.7	170	1	CYHXA	alpha-crystallin c
44	798	87.1	173	1	CYHXA	alpha-crystallin c
45	791	86.4	173	1	CYHXA	alpha-crystallin c

ALIGNMENTS

RESULT 1

alpha-crystallin chain A - human

N/Alternate names: alpha-A-crystallin

C/Species: Homo sapiens (man)

C/Date: 27-Nov-1985 #sequence revision 25-Apr-1997 #text change 22-Jun-1999

C/Accession: S03344; A45947; I56464; A91421; A94588; J03551; I39379; J05690; A02891

R/Jaworski, C.J., Piatigorsky, J.
Nature 337, 752-754, 1989

A/Title: A pseudo-exon in the functional human alpha-A-crystallin gene.
A/Reference number: S03344; MUID:89143747; PMID:2918909

A/Accession: S03344
A/Status: not compared with conceptual translation

A/Molecule type: DNA
A/Residues: 1-104 <JMW>

A/Cross-references: EMBL:X14789; NID:928633; PIDD:CA32891.1; PID:928634
R/MDewitt, D.S.; Hawkins, J.W.; Jaworski, C.J.; Piatigorsky, J.
Exp. Eye Res. 43, 285-291, 1986

A/Title: Isolation and partial characterization of the human alpha-crystallin gene.
A/Reference number: A45947; MUID:87005033; PMID:3758227

A/Accession: A45947
A/Molecule type: DNA

A/Residues: 1-63;166-173 <MCD>
A/Cross-references: GB:M35629; NID:9181077; PIDD:AAA52105.1; PID:9181080; GB:M35628; NI
R/Jaworski, C.J.; Chepelinsky, A.B.; Piatigorsky, J.
J. Mol. Evol. 33, 495-505, 1991

A/Title: The alpha A-crystallin gene: conserved features of the 5'-flanking regions in
A/Reference number: I56464; MUID:92139443; PMID:1779432

A/Accession: I56464
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 1-3 <JMW>

A/Cross-references: GB:S79457; NID:9244474; PIDD:AAA21309.1; PID:9244475
R/de Jong, W.M.; Terwindt, E.C.; Bloemendaal, H.
FEBS Lett. 58, 310-313, 1975

A/Title: The amino acid sequence of the A chain of human alpha-crystallin.
A/Reference number: A91421; MUID:76187952; PMID:817940

A/Accession: A91421
A/Status: compositions of tryptic and thermolytic peptides

A/Molecule type: protein
A/Residues: 132-135;146-151;158-162;166-173 <DBJ>

A/Note: the tryptic peptides were aligned by homology with the bovine and monkey sequence
R/Kramps, J.A.; de Jong, W.M.
submitted to the A148, June 1977

A/Reference number: A94588
A/Accession: A94588

A/Contents: revisions
A/Accession: A94588

A/Molecule type: protein
A/Residues: 1-152;156-173 <KRA>

R/Fujii, N.; Sato, K.; Harada, K.; Ishibashi, Y.
J. Biochem. 116, 663-669, 1994

A/Title: Simultaneous stereoinversion and isomerization at specific aspartic acid residue

A:Reference number: JX0351; MUID:95155281; PMID:7852288
 A:Accession: JX0351
 A:Molecule type: protein
 A:Residues: 1-127,129-173 <FNU>
 A:Experimental source: lens
 A>Note: aspartic acids 58 and 150 are shown to undergo uncatalyzed, aging-related stereoc R;Carpers, G.U.; Penning, J.; De Jong, W.W.
 Exp. Eye Res. 59, 125-126, 1994
 A:Title: A partial cDNA sequence corrects the human alpha A-crystallin primary structure
 A:Reference number: 139379; MUID:95137069; PMID:7835394
 A:Accession: 139379
 A:Status: translated from GE/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 105-156, 'A', 158-160 <CAS>
 A:Cross-references: GB:125781; NID:9531191; PID:AC37570.1; PID:G688439
 R;Roquemore, E.P.; Dell, A.; Morris, H.R.; Panico, M.; Reason, A.J.; Savoy, L.A.; Wistow J. Biol. Chem. 267, 555-563, 1992
 A:Title: Vertebrate lens alpha-crystallins are modified by O-linked N-acetylglucosamine.
 A:Reference number: A58565; MUID:92112709; PMID:1730617
 A:Accession: A58565
 A:Contents: annotation: O-glycosylation
 A>Note: O-glycosylation confirmed but not positioned in human protein
 R;Fuji, N.; Momose, Y.; Yamasaki, M.; Yamagaki, T.; Nakamishi, H.; Uemura, T.; Takita, Biochem. Biophys. Res. Commun. 239, 918-923, 1997
 A:Title: The conformation formed by the domain after alanine-155 induces inversion of as A:Reference number: J05690; MUID:98042494; PMID:9367870
 A:Accession: J05690
 A:Molecule type: protein
 A:Residues: 1-173 <FNU>
 A:Comment: Forms designated A1 and A2 differ by the presence and absence, respectively, C:Comment: This protein is a small heat shock protein and acts as a molecular chaperone C:Genetics:
 A:Gene: GDB:CRYAA; CRYA1
 A:Cross-references: GDB:119074; OMIM:123580
 A:Map position: 21q22.3-21q22.3
 C:Function:
 A:Description: structural component of the eye lens
 C:Superfamily: alpha-crystallin
 C:Keywords: blocked amino end; eye lens; glycoprotein; phosphoprotein
 F:1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental
 F:112/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status predic F:162/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 916; DB 1; Length 173;
 Best Local Similarity 100.0%; Pred. No. 9.2e-81;
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 DB 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 QY 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 DB 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 QY 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173

RESULT 2
 CYEUA
 alpha-crystallin chain A - African elephant (tentative sequence)
 C:Species: Loxodonta africana (African elephant)
 C:Date: 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change 04-Nov-1994
 C:Accession: A02901
 R:De Jong, W.W.; Nuy-Terwindt, E.C.; Versteeg, M.
 Biochim. Biophys. Acta 491, 573-580, 1997
 A:Title: Primary structures of alpha-crystallin A chains of elephant, whale, hyrax and A:Reference number: A90618; MUID:77158093; PMID:870070
 A:Accession: A02901
 A:Molecule type: protein
 A:Residues: 1-173 <DEJ>
 A>Note: compositions of tryptic peptides and sequences of residues 66-67, 69-78, 80, 89-

th the bovine sequence
 C:Superfamily: alpha-crystallin
 C:Keywords: blocked amino end; eye lens
 F:1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match 96.5%; Score 884; DB 1; Length 173;
 Best Local Similarity 95.4%; Pred. No. 1.1e-77;
 Matches 165; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 DB 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 QY 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 DB 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 QY 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173

RESULT 3
 CYRBA
 alpha-crystallin chain A - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 17-Nov-1995
 C:Accession: A02896; B53871
 R:De Jong, W.W.; van der Oudekraak, F.J.; Versteeg, M.; Groenewoud, G.; van Amelsvoort, J.N.
 Eur. J. Biochem. 53, 237-242, 1975
 A:Title: Primary structures of the alpha-crystallin A chains of seven mammalian species.
 A:Reference number: A91230
 A:Accession: A02896
 A:Molecule type: protein
 A:Residues: 1-173 <DEJ>
 A>Note: compositions of tryptic peptides and sequences of residues 2-6 and 146-150 were R;Parveen, R.; Smith, J.B.; Sun, Y.; Smith, D.L.
 J. Protein Chem. 12, 93-101, 1993
 A:Title: Primary structure of rabbit lens alpha-crystallins.
 A:Reference number: A53871; MUID:93151974; PMID:8427639
 A:Accession: B53871
 A:Molecule type: protein
 A:Residues: 1-173 <PAR>
 A:Experimental source: lens
 A>Note: sequence extracted from NCBI backbone (NCBI:124613)
 C:Superfamily: alpha-crystallin
 C:Keywords: acetylated amino end; eye lens; phosphoprotein
 F:1-168/Product: alpha crystallin chain A, minor form #status experimental
 F:112/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 96.4%; Score 883; DB 1; Length 173;
 Best Local Similarity 96.0%; Pred. No. 1.4e-77;
 Matches 166; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 DB 1 MDVTIQHPWFKRTIGPFPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRVLDSG 60
 QY 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 DB 61 ISEVRSDDKVFIFLDVGHFSPEDLTAVKQDDFVEIHGKHNRRQDDHGYSRFFRRYRL 120
 QY 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173
 DB 121 PSNVQASLSCSLSDGMLTFGPGPKIQGLDATHAERAIPIVSRREKPTSPASS 173

RESULT 4
 CYMOA
 alpha-crystallin chain A - rhesus macaque (tentative sequence)
 C:Species: Macaca mulatta (rhesus macaque)

C/Date: 24-Apr-1984 #sequence_revision 27-Nov-1985 #text_change 30-Sep-1993
 C/Accession: A02890
 R/de Jong, W.W.; van der Ouderaa, F.J.; Versteeg, M.; Groenewoud, G.; van Amelsvoort, J.
 Eur. J. Biochem. 53, 237-242, 1975
 A/Title: Primary structures of the alpha-crystallin A chains of seven mammalian species.
 A/Reference number: A91230
 A/Accession: A02890
 A/Molecule type: protein
 A/Residues: 1-173 <DEJ>
 A/Note: Compositions of tryptic peptides and sequences of residues 2-6, 89-96, and 146-1
 C/Superfamily: alpha-crystallin
 C/Keywords: blocked amino end; eye lens
 F/I/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match
 Best Local Similarity 96.0%; Score 879.5; DB 1; Length 173;
 Matches 168; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

QY 121 PSNVDSALSCSLSDAGMLTFPGPKIQTGDATAERAIPVSRREKTSAPSS 173
 121 PSNVDSALSCSLSDAGMLTFPGPKIQTGIDA-HTERAIPVSRREKTSAPSS 172
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

RESULT 5
 alpha-crystallin chain A - guinea pig (tentative sequence)
 C/Species: Cavia porcellus (guinea pig)
 C/Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 31-Dec-1996
 C/Accession: A02894
 R/de Jong, W.W.; Zweers, A.; Goodman, M.
 In Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
 A/Title: Trends in the molecular evolution of alpha-crystallin.
 A/Reference number: A94432
 A/Accession: A02894
 A/Molecule type: protein
 A/Residues: 1-173 <DEJ>
 C/Superfamily: alpha-crystallin
 C/Keywords: acetylated amino end; eye lens
 F/I/Modified site: acetylated amino end (Met) #status predicted

Query Match
 Best Local Similarity 95.9%; Score 878; DB 1; Length 173;
 Matches 165; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

QY 121 PSNVDSALSCSLSDAGMLTFPGPKIQTGDATAERAIPVSRREKTSAPSS 173
 121 PSNVDSALSCSLSDAGMLTFPGPKIQSGDAGHSERAIPVSRREKTSAPSS 173
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

RESULT 6
 alpha-crystallin chain A - springhaas (tentative sequence)
 C/Species: Pedetes capensis (springhaas, springhare)
 C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1996
 C/Accession: E94432; A02894
 R/de Jong, W.W.; Zweers, A.; Goodman, M.
 In Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,

A/Title: Trends in the molecular evolution of alpha-crystallin.
 A/Reference number: A94432
 A/Accession: E94432
 A/Molecule type: protein
 A/Residues: 1-173 <DEJ>
 C/Superfamily: alpha-crystallin
 C/Keywords: acetylated amino end; eye lens
 F/I/Modified site: acetylated amino end (Met) #status predicted

Query Match
 Best Local Similarity 95.9%; Score 878; DB 1; Length 173;
 Matches 165; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

QY 121 PSNVDSALSCSLSDAGMLTFPGPKIQTGDATAERAIPVSRREKTSAPSS 173
 121 PSNVDSALSCSLSDAGMLTFPGPKIQSGDAGHSERAIPVSRREKTSAPSS 173
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

RESULT 7
 alpha-crystallin chain A - brown lemur (tentative sequence)
 C/Species: Lemur fulvus fulvus (brown lemur)
 C/Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 31-Dec-1996
 C/Accession: A02897
 R/de Jong, W.W.; Zweers, A.; Goodman, M.
 In Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
 A/Title: Trends in the molecular evolution of alpha-crystallin.
 A/Reference number: A94432
 A/Accession: A02897
 A/Molecule type: protein
 A/Residues: 1-173 <DEJ>
 C/Superfamily: alpha-crystallin
 C/Keywords: acetylated amino end; eye lens
 F/I/Modified site: acetylated amino end (Met) #status predicted

Query Match
 Best Local Similarity 95.6%; Score 876; DB 1; Length 173;
 Matches 164; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFEXDLLPFLSSTISPYRQSLFRTVLDG 60
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

QY 121 PSNVDSALSCSLSDAGMLTFPGPKIQTGDATAERAIPVSRREKTSAPSS 173
 121 PSNVDSALSCSLSDAGMLTFPGPKIQSGDAGHSERAIPVSRREKTSAPSS 173
 Db 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120
 61 ISEVRSDRDKFVFLDYKHSPEDLTAVKQDDFVEIHGKNERODDHGYISREHRRYRL 120

RESULT 8
 alpha-crystallin chain A - southern American pika (tentative sequence)
 C/Species: Ochotona princeps (southern American pika)
 C/Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 31-Dec-1996
 C/Accession: A02895
 R/de Jong, W.W.; Zweers, A.; Goodman, M.
 In Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
 A/Title: Trends in the molecular evolution of alpha-crystallin.
 A/Reference number: A94432
 A/Accession: A02895
 A/Molecule type: protein
 A/Residues: 1-173 <DEJ>

C:Superfamily: alpha-crystallin
C:Keywords: acetylated amino end; eye lens
F:1/Modified site: acetylated amino end (Met) #status predicted

Query Match 95.3%; Score 873; DB 1; Length 173;
Best Local Similarity 94.8%; Pred. No. 1.3e-76;
Matches 163; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIOHPWFKRTIGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
DB 1 MDVTIOHPWFKRPLGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
QY 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
DB 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173
DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173

RESULT 9

CYLPAA
alpha-crystallin chain A - potto (tentative sequence)

C:Species: Perodicticus potto (potto)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 31-Dec-1996
C:Accession: A02898
R:de Jong, W.W.; Zweers, A.; Goodman, M.
in Protides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
A:Title: Trends in the molecular evolution of alpha-crystallin.
A:Reference number: A94432
A:Accession: A02898
A:Molecule type: protein
A:Residues: 1-173 <DEU>
C:Superfamily: alpha-crystallin
C:Keywords: acetylated amino end; eye lens
F:1/Modified site: acetylated amino end (Met) #status predicted

Query Match 95.3%; Score 873; DB 1; Length 173;
Best Local Similarity 94.2%; Pred. No. 1.3e-76;
Matches 163; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIOHPWFKRTIGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
DB 1 MDVTIOHPWFKRPLGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
QY 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
DB 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173
DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173

RESULT 10

CYGCNA
alpha-crystallin chain A - bush baby (tentative sequence)

C:Species: Galago sp. (bush baby)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1996
C:Accession: P94433; A02898
R:de Jong, W.W.; Zweers, A.; Goodman, M.
in Protides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
A:Title: Trends in the molecular evolution of alpha-crystallin.
A:Reference number: A94432
A:Accession: P94432
A:Molecule type: protein
A:Residues: 1-173 <DEU>
C:Superfamily: alpha-crystallin
C:Keywords: acetylated amino end; eye lens
F:1/Modified site: acetylated amino end (Met) #status predicted

Query Match 95.3%; Score 873; DB 1; Length 173;

Best Local Similarity 94.2%; Pred. No. 1.3e-76;
Matches 163; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIOHPWFKRTIGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
DB 1 MDVTIOHPWFKRPLGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
QY 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
DB 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173
DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173

RESULT 11

CYHXA
alpha-crystallin chain A - Cape rock hyrax (tentative sequence)

C:Species: Procavia capensis (Cape rock hyrax)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 04-Nov-1994
C:Accession: A02902
R:de Jong, W.W.; Nuy-Terwindt, E.C.; Versteeg, M.
Biochim. Biophys. Acta 491, 573-580, 1977
A:Title: Primary structures of alpha-crystallin in chains of elephant, whale, hyrax and ri
A:Reference number: A90618; MUID:77158093; PMID:870070
A:Accession: A02902
A:Molecule type: protein
A:Residues: 1-173 <DEU>
A:Note: compositions of tryptic peptides and sequences of residues 13, 55-57, 66-75, and
C:Superfamily: alpha-crystallin
C:Keywords: blocked amino end; eye lens
F:1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match 95.3%; Score 873; DB 1; Length 173;
Best Local Similarity 93.6%; Pred. No. 1.3e-76;
Matches 162; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDVTIOHPWFKRTIGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
DB 1 MDVTIOHPWFKRPLGPFYPSRLFDQFEGGLFEYDLPLFLSTISPPYRQSLFRTVLDG 60
QY 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
DB 61 ISEVRSDRDKVITFLDVGHFSPEDLTAVKQDDFVEIHGKNERODDHGYISREFFRRYRL 120
QY 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173
DB 121 PSNVDSALSCSLSDAGMLTFGCPKIQTGLDATAERAIPVSRREKPTSPAPSS 173

RESULT 12

CYRTA
alpha-crystallin chain A - rat

C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1982 #sequence_revision 27-Nov-1985 #text_change 16-Jun-2000
C:Accession: A02892; A93739; I55370; I70087
R:de Jong, W.W.; van der Oudekraak, F.J.; Versteeg, M.; Groenewoud, G.; van Amelsvoort, J.N.
Eur. J. Biochem. 53, 237-242, 1975
A:Title: Primary structures of the alpha-crystallin A chains of seven mammalian species.
A:Reference number: A91230
A:Accession: A02892
A:Molecule type: protein
A:Residues: 1-173 <DEU>
A:Note: the compositions of tryptic peptides and sequences of residues 2-6, 13-16, 89-92,
R:Moormann, R.J.M.; van der Velden, H.M.W.; Dodemont, H.J.; Andreoli, P.M.; Bloemendaal, P.
Nucleic Acids Res. 9, 4813-4822, 1981
A:Title: An unusually long non-coding region in rat lens alpha-crystallin messenger RNA.
A:Reference number: A93739; MUID:82081811; PMID:6171772
A:Accession: A93739
A:Molecule type: mRNA

Query Match 95.3%; Score 873; DB 1; Length 173;

A;Residues: 53-173 <MOO>
A;Cross-references: GB:V01219; GB:J00715; NID:955598; PIDN:CAA24530.1; PID:9809074
R;Strinivasan, A.N.; Nagineni, C.N.; Bhat, S.P.
J. Biol. Chem. 267, 23337-23341, 1992
A;Title: Alpha A-crystallin is expressed in non-ocular tissues.
A;Reference number: 155370; MUID:93054670; PMID:1429679
A;Accession: 155370
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 90-172 <RES>
A;Cross-references: GB:M96949; NID:g202619; PIDN:AAA0644.1; PID:g202620
A;Experimental source: spleen
A;Accession: 170087
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 90-123, 'N', 125-172 <RE2>
A;Cross-references: GB:M96950; NID:g202621; PIDN:AAA0645.1; PID:g202622
A;Experimental source: eye
A;Genetics:
A;Gene: alpha A-crystallin
C;Superfamily: alpha-crystallin
C;Keywords: alternative splicing; blocked amino end; eye lens
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match 95.1%; Score 871; DB 1; Length 173;
Best Local Similarity 94.8%; Pred. No. 2e-76;
Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVTIQHPWFKRALGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
QY 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120
DB 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120

QY 121 PSNVDSGALSCLSLADGMLTFGCGPKIQTGDATHAERAIIVSREKETSAPSS 173
DB 121 PSNVDSGALSCLSLADGMLTFGCGPKVQSGLDAGHSERAIIVSREKETSAPSS 173

RESULT 13

CYHAB
alpha-crystallin chain A - Mongolian jird (tentative sequence)
C;Species: Meriones unguiculatus (Mongolian jird)
C;Date: 30-Jun-1988 #sequence_revision 31-Dec-1991 #text_change 24-Nov-1999
C;Accession: C94432; A02892
R;de Jong, W.W.; Zweers, A.; Goodman, M.
in Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
A;Title: Trends in the molecular evolution of alpha-crystallin.
A;Reference number: A94432
A;Accession: C94432
A;Molecule type: protein
A;Residues: 1-173 <DBJ>
C;Superfamily: alpha-crystallin
C;Keywords: alternative splicing; blocked amino end; eye lens
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match 95.1%; Score 871; DB 1; Length 173;
Best Local Similarity 94.8%; Pred. No. 2e-76;
Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVTIQHPWFKRALGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
QY 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120
DB 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120

QY 121 PSNVDSGALSCLSLADGMLTFGCGPKIQTGDATHAERAIIVSREKETSAPSS 173
DB 121 PSNVDSGALSCLSLADGMLTFGCGPKVQSGLDAGHSERAIIVSREKETSAPSS 173

RESULT 14

CYHVA
alpha-crystallin chain A - golden hamster (tentative sequence)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 24-Nov-1999
C;Accession: D94432; A02892
R;de Jong, W.W.; Zweers, A.; Goodman, M.
in Provides of the Biological Fluids, Proc. 28th Colloq., Peeters, H., ed., pp.161-164,
A;Title: Trends in the molecular evolution of alpha-crystallin.
A;Reference number: A94432
A;Accession: D94432
A;Molecule type: protein
A;Residues: 1-173 <DBJ>
C;Superfamily: alpha-crystallin
C;Keywords: alternative splicing; blocked amino end; eye lens
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental

Query Match 95.1%; Score 871; DB 1; Length 173;
Best Local Similarity 94.8%; Pred. No. 2e-76;
Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVTIQHPWFKRALGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
QY 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120
DB 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120

QY 121 PSNVDSGALSCLSLADGMLTFGCGPKIQTGDATHAERAIIVSREKETSAPSS 173
DB 121 PSNVDSGALSCLSLADGMLTFGCGPKVQSGLDAGHSERAIIVSREKETSAPSS 173

RESULT 15

CYMSA
alpha-crystallin chain A - mouse
C;Species: Mus musculus (house mouse)
C;Date: 25-Feb-1985 #sequence_revision 19-Apr-1996 #text_change 22-Jun-1999
C;Accession: A02893; A18860
R;King, C.R.; Shinohara, T.; Platigorsky, J.
Science 215, 985-987, 1982
A;Title: alphaA-crystallin messenger RNA of the mouse lens: more noncoding than coding
A;Reference number: A02893; MUID:83119896; PMID:7156978
A;Accession: A02893
A;Molecule type: mRNA
A;Residues: 11-173 <KT2>
A;Cross-references: GB:J00376; NID:g192760; PIDN:AAA37471.1; PID:g387134
A;Note: the mouse sequence appears to be identical with the rat sequence
R;King, C.R.; Platigorsky, J.
Cell 32, 707-712, 1983

A;Title: Alternative RNA splicing of the murine alphaA-crystallin gene: protein-coding
A;Reference number: A18860; MUID:8315647; PMID:6187470
A;Accession: A18860
A;Molecule type: DNA
A;Residues: 1-104 <KIN>
C;Superfamily: alpha-crystallin
C;Keywords: acetylated amino end; alternative splicing; eye lens
F;1/Modified site: acetylated amino end (Met) #status predicted

Query Match 95.1%; Score 871; DB 1; Length 173;
Best Local Similarity 94.8%; Pred. No. 2e-76;
Matches 164; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDVTIQHPWFKRTLGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
DB 1 MDVTIQHPWFKRALGPPYPSRLFDQFGEGLFYEDLLPFLSSTISPYRQSLFRTVLDG 60
QY 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120
DB 61 ISEVRSRDRKFEVFLDVKHFSPEDLTIVKQDDPVEIHGKNERODDHGYISREPHRRYRL 120

QY 121 PSNVDSGALSCLSLADGMLTFGCGPKIQTGDATHAERAIIVSREKETSAPSS 173
DB 121 PSNVDSGALSCLSLADGMLTFGCGPKVQSGLDAGHSERAIIVSREKETSAPSS 173

Tue Sep 28 12:49:11 2004

us-10-657-740-1.rpr

Page 6

```

QY      121  P S N V D Q S A L S C S L S A D G M L T F C G P K I Q T G L D A T A E R A I P V P S R E E K P T S A P S S 173
      |||||
Db      121  P S N V D Q S A L S C S L S A D G M L T F S G P K V Q S G L D A G H S E R A I P V P S R E E K P S S A P S S 173

```

Search completed: September 27, 2004, 13:21:38
Job time : 16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2004, 12:02:53 ; Search time 77 Seconds
(without alignments)
2681.061 Million cell updates/sec

Title: US-10-657-740-2

Perfect score: 372
Sequence: 1 tccctctccgcacacgtgct.....ctcgctcctcgtcctaa 372

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTus.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.6	31.1	789	4	US-09-023-655-1300 Sequence 1300, App
2	115.6	31.1	1380	4	US-09-513-783A-169 Sequence 169, App
3	114.2	30.7	826	4	US-09-621-976-3172 Sequence 3172, App
4	95	25.5	1379	4	US-09-553-498-5 Sequence 5, Appli
5	95	25.5	1379	4	US-09-618-869-5 Sequence 5, Appli
6	53.8	14.5	495	4	US-09-669-751-69 Sequence 69, Appli
7	53.2	14.3	1627	2	US-08-900-407-2 Sequence 2, Appli
8	43	12.1	44377	2	US-08-804-227C-7 Sequence 7, Appli
9	45	12.1	44377	2	US-08-804-198-1 Sequence 1, Appli
10	43.6	11.7	1929	4	US-09-380-420C-1 Sequence 1, Appli
11	43.6	11.7	1929	4	US-09-899-642A-1 Sequence 1, Appli
12	43	11.6	1820	1	US-08-173-508-7 Sequence 7, Appli
13	43	11.6	1821	2	US-08-265-310-7 Sequence 7, Appli
14	43	11.6	1821	2	US-08-951-742-7 Sequence 7, Appli
15	42.6	11.5	3624	1	US-07-951-715A-6 Sequence 6, Appli
16	42.6	11.5	3624	2	US-08-459-448A-6 Sequence 6, Appli
17	42.6	11.5	3624	2	US-08-459-595A-6 Sequence 6, Appli
18	42.6	11.5	3624	3	US-08-459-504B-6 Sequence 6, Appli
19	42.6	11.5	3624	3	US-08-459-444-6 Sequence 6, Appli
20	42.6	11.5	3624	3	US-09-053-549-7 Sequence 7, Appli
21	42.6	11.5	3624	4	US-09-547-422-6 Sequence 6, Appli
22	42.2	11.3	1008	4	US-09-199-637A-148 Sequence 148, App
23	42.2	11.3	3156	4	US-09-252-991A-11652 Sequence 11652, A
24	42	11.3	1536	4	US-09-252-991A-11704 Sequence 11704, A
25	42	11.3	1761	4	US-09-252-991A-11782 Sequence 11782, A
26	42	11.3	2118	4	US-09-252-991A-11782 Sequence 11782, A
27	41.8	11.2	1317	4	US-09-252-991A-779 Sequence 779, App

28	41.8	11.2	1407	4	US-09-252-991A-637 Sequence 637, App
29	41.8	11.2	1728	4	US-09-252-991A-616 Sequence 616, App
30	41.2	11.1	1338	2	US-08-044-812A-3 Sequence 3, Appli
31	41.2	11.1	1338	2	US-08-475-637-3 Sequence 3, Appli
32	41.2	11.1	1338	3	US-08-706-281A-11 Sequence 11, Appli
33	41.2	11.1	1338	3	US-09-191-359-3 Sequence 3, Appli
34	41.2	11.1	1338	3	US-09-097-231-11 Sequence 11, Appli
35	41.2	11.1	1338	4	US-09-353-099-11 Sequence 11, Appli
36	41	11.0	1428	4	US-09-252-991A-2575 Sequence 2575, App
37	41	11.0	1725	4	US-09-252-991A-2341 Sequence 2341, App
38	40.8	11.0	43280	2	US-08-804-227C-1 Sequence 1, Appli
39	40.4	10.9	309	4	US-09-252-991A-15583 Sequence 15583, A
40	40.4	10.9	1413	4	US-09-252-991A-15729 Sequence 15729, A
41	40.4	10.9	1527	4	US-09-252-991A-15698 Sequence 15698, A
42	40.4	10.9	1791	4	US-09-252-991A-15692 Sequence 15692, A
43	40.2	10.8	1068	4	US-09-252-991A-9810 Sequence 9810, App
44	40.2	10.8	1092	4	US-09-252-991A-10170 Sequence 10170, A
45	40.2	10.8	2472	4	US-09-252-991A-886 Sequence 886, App

ALIGNMENTS

RESULT 1
US-09-023-655-1300
Sequence 1300, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
TITLE OF INVENTION: Jeffrey J. Seilhammer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023.655
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 845-4166
TELEFAX: (650) 845-0555
INFORMATION FOR SEQ ID NO: 1300:
SEQUENCE CHARACTERISTICS:
LENGTH: 789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g32477
US-09-023-655-1300
Query Match 31.1%; Score 115.6; DB 4; Length 789;

Best Local Similarity 62.1%; Pred. No. 1e-20; Indels 3; Gaps 1;
Matches 200; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 27 CGGCACTCTGAGGTTGATCCGACCGGAGCAAGTTCTGATCTCTGATGTAAGCA 86
DB 301 CGGGGTCTCGAGATCCGACACTGCGGACCGCTGGCGCGTCTCTGATGTAAGCA 360

QY 87 CTTTCCCGGAGAGCTTACCGTGAAGGTGACGACGACTTTGAGAGATCCAGGAAA 146
DB 361 CTTCCGCCCCGACGACGTGACCGTCAAGACCAAGATGCGGAGTGAAGATCACCAGCAA 420

QY 147 GCACGAGAGGCGCCGACGACGACGCTACATTTCCGAGTCCACCGCGCTACCG 206
DB 421 GCACGAGAGGCGCCGACGACGACGCTACATTTCCGAGTCCACCGCGCTACCG 480

QY 207 CCGCGCTGCAAGCTGACGACGACGCTCTCTGCTCTCTGCTGCGGATGCGATGCT 266
DB 481 GGTGCCCCCGGTGTGAGACCCACCAAGTTCTCTCTCTGCTGCGGAGGACACT 540

QY 267 GACCTTCTGTGGCCCCCAAGATCAGACTGGCGCTGATGCCACCA---CGCCGAGCGAGC 323
DB 541 GACCGTGAAGGCCCCCATGCCCCAAGCTAGCCACGACGATCCAGATCACCATCCAGT 600

QY 324 CATCCCGGTGCGGGGAGAG 345
DB 601 CACCTTGAGTGGCGGCCAG 622

RESULT 2

US-09-513-783A-169
; Sequence 169, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliani, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-11
; CURRENT APPLICATION NUMBER: US/09/513,783A
; NUMBER OF SEQ ID NOS: 2000-02-25
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 169
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
; NAME/KEY: CDS
; LOCATION: (1)..(1380)
US-09-513-783A-169

Query Match 31.1%; Score 115.6; DB 4; Length 1380;
Best Local Similarity 62.1%; Pred. No. 1.2e-20;
Matches 200; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 27 CGGCACTCTGAGGTTGATCCGACCGGAGCAAGTTCTGATCTCTGATGTAAGCA 86
DB 1011 CGGGGTCTCGAGATCCGACACTGCGGACCGCTGGCGCGTCTCTGATGTAAGCA 1070

QY 87 CTTTCCCGGAGAGCTTACCGTGAAGGTGACGACGACTTTGAGAGATCCAGGAAA 146
DB 1071 CTTGCCCCCGGACGACGTGACCGTCAAGACCAAGATGCGGAGTGAAGATCACCAGCAA 1130

QY 147 GCACGAGAGGCGCCGACGACGACGCTACATTTCCGAGTCCACCGCGCTACCG 206
DB 1131 GCACGAGAGGCGCCGACGACGACGCTACATTTCCGAGTCCACCGCGCTACCG 1190

QY 207 CCGCGCTGCAAGCTGACGACGACGCTCTCTGCTCTCTGCTGCGGATGCGATGCT 266
DB 1191 GGTGCCCCCGGTGTGAGACCCACCAAGTTCTCTCTCTGCTGCGGAGGACACT 1250

QY 267 GACCTTCTGTGGCCCCCAAGATCAGACTGGCGCTGATGCCACCA---CGCCGAGCGAGC 323
DB 601 CACCTTGAGTGGCGGCCAG 622

DB 1251 GACCTGAGAGGCCCCCATGCCCCAAGCTAGCCACGATCTCAAGAGATCACCATCCAGT 1310
QY 324 CATCCCGGTGTCGCGGAGAG 345
DB 1311 CACCTTGAGTGGCGGCCAG 1332

RESULT 3

US-09-621-976-3172
; Sequence 3172, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jodert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3172
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 557..736
US-09-621-976-3172

Query Match 30.7%; Score 114.2; DB 4; Length 826;
Best Local Similarity 65.0%; Pred. No. 2.3e-20;
Matches 184; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 10 CGCACCGTGTGAGTCCGACATCTGTGAGTTGATCCGACCGGAGCAAGTTCTGATC 69
DB 442 CCGACCTGTTGAGTACGACTGACGACTGACGAGATGGCGCGGAGAGACAGGTTCTGTGC 501

QY 70 TTCTCGATGAGACACTTCTCCCGGAGAGCTTCAACCGTGAAGTGCAGGACACTTT 129
DB 502 AACCTGATGTGAGACACTTCTCCCGGAGAGCAAGTCAAGTGAAGTGTGGAGATGTC 561

QY 130 G-TGGAGATCCACGGAAGCAACAGGAGCCAGGACGACGCTCATTTCCGTTA 188
DB 562 GATTAGGTGATGGAAGAAACATGAAGAGGCGGAGTGAATGATTTTCATCTCCAGGGA 621

QY 189 GTTCACCGCGCTACCGCTGCGCTGCAACGTGACGAGTGGGCCCTCTTGTCTCCCT 248
DB 622 GTTCACAGGAATATCCGAGATCCAGCTGATGATGACCTTCACCATTAATTCAATCCCT 681

QY 249 GTTCCCGATGAGTGTGACCTTCTGTGAGGCCCCCAAGATCCAG 291
DB 682 GTCATCTGATGGGATCTTCATCTGATGATGACCAAGAAACAG 724

RESULT 4

US-09-553-498-5
; Sequence 5, Application US/09553498
; Patent No. 6309861
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorothee
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: Process for the production of naturally folded and secreted protein
; FILE REFERENCE: Case 20379
; CURRENT APPLICATION NUMBER: US/09/553,498
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: EP99107412.1
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 5
; LENGTH: 1379


```

: TYPE: DNA
: ORGANISM: E. coli
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (392) .. (1090)
US-09-553-498-5

```

Query Match	25.5%;	Score 95;	DB 4;	Length 1379;
Best Local Similarity	60.8%;	Pred. No. 2.1e-15;		
Matches 155; Conservative	0;	Mismatches 100;	Indels 0;	Gaps 0;

[illegible]

RESULT 5
US-09-618-869-5

```

? Sequence 5: Application US/03618869
? Patent No. 6455279
? GENERAL INFORMATION:
? APPLICANT: Ambrosius, Dorothee
? APPLICANT: Rudolph, Rainer
? APPLICANT: Schaeffner, Joerg
? APPLICANT: Schwarz, Elisabeth
? TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND
? TITLE OF INVENTION: SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR
? FILE REFERENCE: 20381
? CURRENT APPLICATION NUMBER: US/09/618,869
? CURRENT FILING DATE: 2000-07-19
? PRIOR APPLICATION NUMBER: EP99114811.5
? PRIOR FILING DATE: 1999-07-29
? NUMBER OF SEQ ID NOS: 10
? SOFTWARE: SeqEditor Ver. 2.1
? SEQ ID NO 5
? LENGTH: 1379
? TYPE: DNA
? ORGANISM: Escherichia coli
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (382)..(1090)
? OS-09-618-869-5

```

Query Match	25.5%	Score 95;	DB 4;	Length 1379;
Best Local Similarity	60.8%	Pred. No. 2.1e-15;		
Matches 155;	Conservative 0;	Mismatches 100;	Indels 0;	Gaps 0;
QY	18	GCTGACTCCGGCACTCTGAGTTCATCCGACCGGAGCAAGTTCGTCATCTTCCTGGA	77	
Db	712	GCTCAGAGAGCGGGTCTCGAGATCCAGACAGCGCTGATGCGCGGTCTCCCTGGA	771	
QY	78	TGTGAAGCACTTCTCCCGGAGAGCACTCACCGTGAAGGTGAGAGCACTTGTGAGAT	137	
Db	772	CGTCAACCACTTCGCTCCGGAGAGGCTCAGGTGAAGCAAGAGAGCGGTGTGAGAT	831	
QY	138	CCAGGAAAGCACACAGACGCGCCAGAGACCAAGGCTACTTTCCTGATTCACCG	197	

[illegible]

RESULT 6
US-09-669-751-69

```

Sequence: 697 Application US/09669/751
Patent No. 6551575
GENERAL INFORMATION:
APPLICANT: Greenspan, Ralph J.
TITLE OF INVENTION: Methods for Identifying Compounds for
TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
TITLE OF INVENTION: Balance and the Perception of Gravity
FILE REFERENCE: P-NI 3864
CURRENT APPLICATION NUMBER: US/09/669,751
CURRENT FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US 60/168,579
PRIOR FILING DATE: 1999-12-02
NUMBER OF SEQ ID NOS: 261
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 69
LENGTH: 495
TYPE: DNA
ORGANISM: Drosophila
US-09-669-751-69

```

Query Match	14.5%	Score 53.8	DB 4	Length 495
Best Local Similarity	67.3%	Pred. No. 5.5e-05		
Matches 76	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

QY 61 TTTCGATCTTCTCGATGTAAGACCTTCCCCGAGAGACCTCACCGTGAAGGTGAG 120
Db 377 TTTCAGGTGTGATGGATGTGTCTCGACTTCAAGCCCAAGACTGACCGTCAAGGTGTG 436	
QY 121 GACGACTTTGTGAGATCCACGGAAGCAACAAGAGGCCGAGACGACCAAG 173	
Db 437 GACCAACCGTGTGTGTAAGAGGCAAGACCAAGAGAGCGCCACTTG 489	

RESULT 7
US-08-900-407-2

Sequence 2, Application US/08900407
Patent No. 5962262

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Puryi
TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTSEQ for Windows Version 2.0
CURRENT APPLICATION NUMBER: US/08/900,407
APPLICATION NUMBER: US/08/900,407

APPLICATION NUMBER: US/08/900,407

FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0351 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1627 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE: 1
INDIVIDUAL ISOLATE:
IMMEDIATE SOURCE:
LIBRARY: 222?
CLONE: 1362715
US-08-900-407-2

Query Match 14.3%; Score 53.2; DB 2; Length 1627;
Best Local Similarity 52.8%; Pred. No. 0.00011;
Matches 115; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 77 ATGTGAACACTCTCCCGGAGACCTCACCGTGAGTGAGAGCACTTTGTGAGA 136
DB 818 ATGTGAACACTCTCAAGCCAGAGAGTTGATGTATGAAGCAAGATGATGAGAGG 877
QY 137 TCCAGGAAGACACAGAGCGCCAGAGACACCGCTACATTTCCGTTGATCCAC 196
DB 878 TGTCTGGAAACATAGAGAAAGCAAGAGAGTGAGTTGTTCTAAGAACTTCACAA 937
QY 197 GCCGTACCGCTGCGCTTCCACAGTGAACAGTGGCCCTCTCTGCTTCTGCTGCGG 256
DB 938 AGAAATCCAGCTTCTCTGAGAGTGAGATCTGTACAGTATTTGCTCATTTCCAG 997
QY 257 ATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGACT 294
DB 998 AGGTCTGTGATCATCGAAGCTCCCGAGTCTCTCT 1035

RESULT 8
US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 12.1%; Score 45; DB 2; Length 44377;
Best Local Similarity 46.4%; Pred. No. 0.031;
Matches 147; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 12 CACGCTGTGACTTCGCGATCTGTGAGTTGATCGACCGGAGCAAGTTGTCATCTT 71
DB 30078 CACGCTGTGCTCTGCGACGCTCTCGACGCGGACCGCTCGCGAGCTGCGCGCAT 30137
QY 72 CCTGATGTGAAGCACTTCTCCCGAGAGACTTCAACCGTGAAGTGCAGAGCACTTTGT 131
DB 30138 ACCGCGCATGACCCGCTGAGAGCGCGGTGTGACAGCACCGGCTCTCGAGAGCGGT 30197
QY 132 GAGATCCAGGAAAGCAACAAGAGCGCCAGAGCGAGCAAGCTTCCCGTGAATT 191
DB 30198 GCTGCGCGGGCTCACCCCGAGCGAGATGCGCGCTGTGCGGCCCAAGTGAAGCGCG 30257
QY 192 CCACGCGCTACCGCGCTGCGCTCAACGTGACCAAGTGGCCCTCTGCTGCTCTGTC 251
DB 30258 GGTCCACTGAGCAACTCACCCGCGACCTCGACTGTGCGGTCTCTTCTCTCTC 30317
QY 252 TGCCGATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGACTGAGCTGATGCCACCA 311
DB 30318 CAGCGCGGTCTGTGCGAGCGCGCCCAAGGCAACTAGCGCGGCCAAGCCACCT 30377
QY 312 CGCCGAGCGAGCCATCC 328
DB 30378 CGAGCGCTCGCGCGCC 30394

RESULT 9
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match 12.1%; Score 45; DB 2; Length 44377;
Best Local Similarity 46.4%; Pred. No. 0.031;
Matches 147; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 12 CACCGTCTGGACCTCCGGCATCTCTGAGTTCGATCCGACCGGACAAAGTTCTGATCTT 71
DB 30078 CACCGTGTGCTGCTGCGACGCTCTCGACGCGGACGCGCTCGGACTGCTCGCGCAT 30137

QY 72 CCTCGATGTGAAGCACTTCTCCCGGAGACCTCACCGTGAAGGTGAGAGCACTTGT 131
DB 30138 ACCGCGCATACCCGCTGACGCGGTGTGTGACAGACGCGGCTCTCGACGAGCGGT 30197

QY 132 GGAAGATCACGAAAGACAAAGACGCGCAGAGACGACCAAGCTATTTCCGTGAGTT 191
DB 30198 GCTGCGCGGCTACCCCGAGCGGATGCGGCGGTGCTGCGGCCCAAGGTGAGAGCGCG 30257

QY 192 CCACCGCGCTACCGCTTGCCTGCTCAACGTGACCAATCGGCTCTCTTGTCTCTGTC 251
DB 30258 CGTCCACTGAGACCACTACCCCGGACCTCGACTGTGCGCGTTCTCTCTCTCTC 30317

QY 252 TGGCGATGCGATGCTGCTTCTGTGACCCCAAGATCCAGACTGAGCTGATGACCA 311
DB 30318 CAGGCGCGGTCTGTGTGAGGACGCGGCTCAGGGAACCTACGCGGCGCAAGCTCACT 30377

QY 312 CGCCGAGCGAGCCATCC 328
DB 30378 CGACGCGCTCGCGCGCC 30394

RESULT 10

US-09-380-420C-1
Sequence 1, Application US/09380420C
Patent No. 6300544
GENERAL INFORMATION:
APPLICANT: Halkier, Barbara
Bak, Soren
Kahn, Rachel
Moller, Birger
TITLE OF INVENTION: Cytochrome P450 Monooxygenases
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: SynGene Patent Dept.
STREET: 3054 Cornwallis Road
CITY: RTP
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/380,420C
FILING DATE: 12-No. 6300544-1999
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-21251A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1929 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: P450ox
FEATURE:
NAME/KEY: CDS
LOCATION: 81..1673
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-380-420C-1

Query Match 11.7%; Score 43.6; DB 4; Length 1929;
Best Local Similarity 49.6%; Pred. No. 0.032;
Matches 112; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1 TCCTCTTCCGACCGGTGCTGACTCCGGATCTGTAGGTTGATCCGACCGGACAA 60
DB 651 TCCATGCTCTCAACGACGCTTCTGCTTCAACGACGATCATCGGACCGTGGCG 710

QY 61 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGACCTCACCGTGAAGTGCAG 120
DB 711 TTGGGAACATCTACGCTTCCAGCACTTGGCCGACAAAGAGCGCTTCCAGACGTGCTG 770

QY 121 GACGACTTTGTGAGATTCACGGAAGACACAGAGCGCCAGGACGACCAAGCTACATT 180
DB 771 GAGAGCGCATGTGACATGATGAGCCAGCTTCTCCGCGAGGACTTCTTCCCAAGCGCGCG 830

QY 181 TCCCGTGAAGTTTCACGCGCGCTACCGCTGCGCTGCTTCAAGCTGAGCC 226
DB 831 GGCCTGCTCGCGACGCTCTGCGGCTTCTGCGCGCGCGGAGC 876

RESULT 11
US-09-899-642A-1
Sequence 1, Application US/09899642A
Patent No. 6649814
GENERAL INFORMATION:

Db 864 GCGGACCTACGCG 878

RESULT 13

US-08-265-310-7
Sequence 7, Application US/08265310
Patent No. 5856166
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
APPLICANT: Malek, Lawrence T.
APPLICANT: Soostmeyer, Gisela
APPLICANT: Walczyk, Eva
APPLICANT: Krygsmann, Phyllis
TITLE OF INVENTION: STREPTOMYCES PROTEINASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,310
FILING DATE: 24-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,508
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/133/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 104..1720
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..244
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 245..1720
US-08-265-310-7

Query Match 11.6%; Score 43; DB 2; Length 1821;

Best Local Similarity 46.0%; Pred. No. 0.044;

Matches 145; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 2 CCCTCTCCGACCGCTGCTGACCTCCGACATCTGTGAGTTTCATCCGACCGGAGCAAGT 61
DB 564 CCGAGTACGACATGCTGCGGTGACCCCGGCGGCGTGGCCCGACGTGAACCCGTCGAGT 623

QY 62 TCGTCATCTTCTCTGATGTGAAGCATTTCCCGGAGCACTTCACCGTGAAGTGCAG 121
DB 624 GCGTGAACGGGCGCGATGATGAGACCGCTACACGCGACCGACGATCACCCCGGACGACGGG 683
QY 122 ACGACTTTGTGAGATCCACGAAAAGCAACAGAGCGCCAGGACGACCGCTACATTT 181
DB 684 GCGAGACGACGAGCTGCTGACGCTTACAAAGAGTTGCGCGAGGGCTGCGGGCGAG 743
QY 182 CCGGTGATTCACCGCGCTACCGCTGCCGTCCACAGTGGACCACTGCGGCTCTCTT 241
DB 744 CGCGAAGCTGCTGCGGCTACGCTTCACGGTGCAGGCGACGCGACATGAGCGTCTGC 803
QY 242 GCTCCGTCTGCGCGATGATGATGATCTTCTGTGCGCCCAAGATCCAGACTGCGCTTG 301
DB 804 GCGGCTGCTGCGGCGAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 863
QY 302 ATGCGACCGACGCGG 316
DB 864 GCGGACCTACGCG 878

RESULT 14

US-08-951-742-7
Sequence 7, Application US/08951742
Patent No. 6127144
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Michael J. Butler
APPLICANT: Dany Hadary
APPLICANT: David Jenish
APPLICANT: Tim Krieger
APPLICANT: Lawrence T. Malek
APPLICANT: Gisela Soostmeyer
APPLICANT: Eva Walczyk
APPLICANT: Phyllis Krygsmann
APPLICANT: Sheila Garven
TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
TITLE OF INVENTION: BACTERIAL HOST CELLS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,742
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 0189740/0140
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 104..1720

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..244
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 245..1720
US-08-951-742-7

Query Match 11.6%; Score 43; DB 3; Length 1821;
Best Local Similarity 46.0%; Pred. No. 0.044;
Matches 145; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 2 CCGCTTCCGACCGCTGAGACTCCGCGATCTCTGAGATTGATCCACCGGACAACT 61
DB 564 CCGGATGACATGATGAGGCTGACCCCGGGGCTGCGGAGTAAACCCGTCGAGT 623
QY 62 TCGTCATCTTCTCGATGTAAGCACTTCTCCCGGAGACCTCAACCGTGAAGTGACAG 121
DB 624 GCTTGAGAGGCGGAGATGACGCTACAGCGGACCGACGTCACCCCGGACGACGCG 683
QY 122 ACGACTTGTGAGATCCACGGAAGCAACAGAGCGGACGACGACGCTTACATTT 181
DB 684 GCGAGACGAGAGAGCTGCTGACGCGCTTCAAGAGATTGCGCGAGGGCTGCGGGGCGAG 743
QY 182 CCGTGATTCACCGCGGCTACCGCTGCGCTGCAAGTGAACCACTGCGCTCTCTT 241
DB 744 CGCCGAACTCTGCGGACGCTTCTCAAGTGTGAGGCGGACGCGACATGACGCTCTGC 803
QY 242 GCTCCCTGTGCGGATGAGCACTGCTGCTGTGTGCGCCCAAGATCCAGACTGACCTG 301
DB 804 GCGGCTGCTGAGGACGAGAGCTGACCTGATGAGGAGCGCTGACGCGACCTCTCTG 863
QY 302 ATGCCACCCACGCG 316
DB 864 GCGGACCTACGCG 878

RESULT 15

US-07-951-715A-6

Sequence 6, Application US/07951715A

Patent No. 5625136

GENERAL INFORMATION:

APPLICANT: Kozziel, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Ewola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Meriin, Ellis J.
APPLICANT: Launis, Karen J.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttle, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/951,715A
FILING DATE: 25-SEP-1992
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 3624 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:

NAME/KEY: CDS
LOCATION: 1..3621
OTHER INFORMATION: /product= "Full-length, maize
optimized cry1B"
OTHER INFORMATION: /note= "Disclosed in Figure 6."

US-07-951-715A-6

Query Match 11.5%; Score 42.6; DB 1; Length 3624;
Best Local Similarity 51.3%; Pred. No. 0.067;
Matches 99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 92 CCGGAGAGACCTCAACGCTGAAGTGACAGAGCACTTGTGAGATCCACGGAAGCA 151
DB 1235 CTTAGAGAGCCCGGCTGTGAGTGAAGAGACGAGACGAGTGCCTCCGAGACCA 1294
QY 152 ACGAGCGGCAAGAGACCAACGCTACATTTCCGTTCCACCGCGCTACCGCTGC 211
DB 1295 CCGAGCGCCCAACTACAGAGACTACAGCCACCGCTGAGCCACATCGGCATCTCTGC 1354
QY 212 CGTCAAGTGAACAGTGGCCCTCTTGTGCTCCCTGTCTGCGCATGCTGACCT 271
DB 1355 AGAGCGGCTGAACGTGCCGCTGTACAGCTGACCCACCGAGCGGACGACCAACA 1414
QY 272 TCTGTGCCCCCA 284
DB 1415 CCATCGGCCCA 1427

Search completed: September 27, 2004, 13:35:28
Job time : 78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2004, 12:02:53 ; Search time 3844 Seconds

(without alignments)
490.156 Million cell updates/sec

Title: US-10-657-740-2

Perfect score: 372

Sequence: 1 tccctcttcgcacacgtgctc.....cctcgctccctcgctcctaa 372

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3337386 seqs, 2532474682 residues

Total number of hits satisfying chosen parameters: 6674772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	372	100.0	372	US-10-657-740-2	Sequence 2, Appl1
2	315.2	84.7	531	US-10-105-427-1	Sequence 1, Appl1
3	295.2	79.4	1056	US-10-316-253-112	Sequence 112, App
4	263	70.7	1271	US-10-316-253-110	Sequence 110, App
5	215.4	57.9	573	US-10-029-386-12881	Sequence 12881, A
6	211	56.7	211	US-10-029-386-26581	Sequence 26581, A
7	129.6	34.8	528	US-10-152-3194-1574	Sequence 1574, Ap
8	129.6	34.8	1247	US-09-917-800A-1419	Sequence 1419, Ap
9	123.6	33.2	691	US-09-954-456-514	Sequence 514, Ap
10	123.6	33.2	691	US-09-960-706-869	Sequence 869, App
11	123.6	33.2	691	US-09-873-319-566	Sequence 566, App
12	123.6	33.2	691	US-09-873-367C-764	Sequence 764, App
13	123.6	33.2	691	US-10-342-887-686	Sequence 686, App
14	123.6	33.2	691	US-10-172-118-686	Sequence 686, App

15	123.6	33.2	691	US-10-133-937-61	Sequence 61, Appl
16	123.6	33.2	691	US-10-159-563-61	Sequence 61, Appl
17	123.2	33.1	548	US-10-101-510-250	Sequence 250, App
18	117	31.5	380	US-09-960-352-4143	Sequence 4143, Ap
19	116.2	31.2	449	US-09-960-352-4277	Sequence 4277, Ap
20	115.6	31.1	599	US-10-060-036-43	Sequence 43, Appl
21	115.6	31.1	764	US-10-605-498-91	Sequence 91, Appl
22	115.6	31.1	789	US-10-641-643-1300	Sequence 1300, Ap
23	115.6	31.1	847	US-10-153-668-283	Sequence 283, App
24	115.6	31.1	865	US-09-969-034-4480	Sequence 4480, App
25	115.6	31.1	865	US-10-342-887-626	Sequence 626, App
26	115.6	31.1	865	US-10-172-118-626	Sequence 626, App
27	115.6	31.1	1231	US-09-880-107-3865	Sequence 3865, App
28	115.6	31.1	1380	US-10-100-957A-169	Sequence 169, App
29	111.2	29.9	708	US-09-814-353-14770	Sequence 14770, A
30	110.4	29.7	954	US-09-814-353-20273	Sequence 20273, A
31	110.2	29.6	614	US-09-814-353-2040	Sequence 2040, Ap
32	110.2	29.6	614	US-09-814-353-836	Sequence 836, Ap
33	110	29.6	500	US-09-918-995-22134	Sequence 22134, A
34	106.8	28.7	1310	US-10-152-319A-1412	Sequence 1412, Ap
35	106.8	28.7	1700	US-09-813-358-2	Sequence 2, Appl
36	106.8	28.7	1700	US-09-997-279-2	Sequence 2, Appl
37	103.6	27.8	373	US-10-085-783A-26398	Sequence 26398, A
38	103.6	27.8	373	US-10-242-535A-26398	Sequence 26398, A
39	103.4	27.8	349	US-09-990-747-28	Sequence 28, Appl
40	102.8	27.6	450	US-09-918-995-4949	Sequence 4949, Ap
41	102.6	27.6	366	US-09-960-352-12622	Sequence 12622, A
42	101.2	27.2	503	US-09-911-904-146	Sequence 146, App
43	100.4	27.0	405	US-09-918-995-4075	Sequence 4075, Ap
44	99.2	26.7	563	US-10-198-846-8376	Sequence 8376, Ap
45	98.6	26.5	411	US-10-198-846-10360	Sequence 10360, A

ALIGNMENTS

RESULT 1

US-10-657-740-2

Sequence 2, Appl

Publication No. US20040157289A1

GENERAL INFORMATION:

APPLICANT: Salemo, John C.

APPLICANT: Hanna, Michael

APPLICANT: Koretz, Jane F.

APPLICANT: Crone, Donna

APPLICANT: Smith, Susan E.

TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM

FILE REFERENCE: 01794100406US1

CURRENT APPLICATION NUMBER: US/10/657,740

CURRENT FILING DATE: 2003-09-08

PRIOR APPLICATION NUMBER: US 60/408,680

PRIOR FILING DATE: 2002-09-06

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 372

TYPE: DNA

ORGANISM: Homo sapiens

US-10-657-740-2

Query Match 100.0%; Score 372; DB 17; Length 372;

Best Local Similarity 100.0%; Pred. No. 1.5e-99;

Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCCTCTTCGCGACCGTCTGACTCGGCATCTTGAGTTGATTCGACCGGACAG 60

1 TCCCTCTTCGCGACCGTCTGACTCGGCATCTTGAGTTGATTCGACCGGACAG 60

Db 1 TCCCTCTTCGCGACCGTCTGACTCGGCATCTTGAGTTGATTCGACCGGACAG 60

QY 61 TTGGTATCTTCCTCCATGGAAGCACTTCCCGGAGGACCTCCGGAAGGTGAG 120

61 TTGGTATCTTCCTCCATGGAAGCACTTCCCGGAGGACCTCCGGAAGGTGAG 120

Db 61 TTGGTATCTTCCTCCATGGAAGCACTTCCCGGAGGACCTCCGGAAGGTGAG 120

QY 121 GACGACTTGTGAGATCCGCGAAGCAACGAGCGCGAGACGACGAGCTACATT 180

QY	Db
181	121
181	181
181	181
241	241
241	241
301	301
301	301
361	361
361	361

```

RESULT 2
US-10-105-427-1
: Sequence 1, Application US/10105427
: Publication No. US20020177192A1
: GENERAL INFORMATION:
: APPLICANT: Council of Scientific and Industrial Research
: TITLE OF INVENTION: Chimeric protein alpha BNAC crystallin with extraordinarily high
: FILE REFERENCE: US- 649
: CURRENT APPLICATION NUMBER: US/10/105,427
: CURRENT FILING DATE: 2002-07-03
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 531
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: DNA sequence for chimeric alpha BNAC
US-10-105-427-1

```

Query Match	Similarity	Score	DB	Length
Best Local	92.2%	315.27	DB 14	531
Matches	332	Conservative	0	Mismatches 28; Indels 0; Gaps 0

QY	10	CGACCGTGTGACTCCGGACATCTGTGAGTTTCGATCCGACCGGGACAAGTTGCTATC	69
Db	172	CCCACTGGTTTGACACTGACCTCTACAGATAGCGCTTGGAAAGGACAGTTCTCTGTC	231
QY	70	TTCTCTGATGTGAACACTTCTCCCGGAGACCTCACTGTAAGTGTGAGACGACTT	129
Db	232	AACTGGATGTGAACACTTCTCCCGGAGACCTCACTGTAAGTGTGAGACGACTT	291
QY	130	GTGGAGATTCACCGAAGCACAACAGCGCCACGAGACGACCAAGCTATTTCCCGTAG	189
Db	292	GTGAGATTCACCGAAGCACAACAGCGCCACGAGACGACCAAGCTATTTCCCGTAG	351
QY	190	TTCCACCGCGCTACCGCTCGCTGCCATCGTGAACGATGGCGCCCTCTCTTGCTCCTG	249
Db	352	TTCCACCGCGCTACCGCTCGCTGCCATCGTGAACGATGGCGCCCTCTCTTGCTCCTG	411
QY	250	TTGTGCGGATGCGATGCTGACCTTCTGTGGCCCCCAATTCACAGCTGGCTGGATGCGATC	309
Db	412	TTGTGCGGATGCGATGCTGACCTTCTGTGGCCCCCAATTCACAGCTGGCTGGATGCGATC	471
QY	310	CACGCGGAGCGAGCATCCCGGTGTGCGGGAGGAAGACCCACCTCGGCTCCCTGTGC	369
Db	472	CACGCGGAGCGAGCATCCCGGTGTGCGGGAGGAAGACCCACCTCGGCTCCCTGTGC	531

RESULT 3
US-10-316-253-112
; Sequence 112, Application US/10316253

```

? Publication No. US20030162706A1
? GENERAL INFORMATION:
? APPLICANT: The Procter & Gamble Company
? APPLICANT: Peters, Kevin
? APPLICANT: Thompson, Kevin
? APPLICANT: Wang, Feng
? APPLICANT: Greis, Kenneth
? TITLE OF INVENTION: Angiogenesis Modulating Proteins
? FILE REFERENCE: 8865M
? CURRENT APPLICATION NUMBER: US/10/316,253
? CURRENT FILING DATE: 2002-12-10
? PRIOR APPLICATION NUMBER: US 60/355,295
? PRIOR FILING DATE: 2002-02-08
? NUMBER OF SEQ ID NOS: 308
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 112
? LENGTH: 1056
? TYPE: DNA
? ORGANISM: Rattus norvegicus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (13)..(534)
? OTHER INFORMATION:
US-10-316-253-112

```

Query Match	79.4%;	Score 295.2;	DB 15;	Length 1056;
Best Local Similarity	87.1%;	Pred. No. 6.2e-77;		
Matches 324;	Conservative 0;	Mismatches 48;	Indels 0;	Gaps 0;
QY	1	TCCTCTTTCGCACCCGCTGCTGACTCCGGCACTCTGTAGGTTTCATCCGACCCGGACAAAG	60	
Db	163	TCCTCTTTCGCACAAAGTTTGAGACTCCGGCACTCTGTAGGTTTCATCCGACCCGGACAAAG	222	
QY	61	TTGCTCATCTTCCCGCATGTGTAAGACATCTTCCCGAGAGACCTACCGTGAAGGTGACG	120	
Db	223	TTTGTCATCTTCTTGAGATGTAAAGCACTTCTCTGTAGGACTTACCGTGAAGGTACTCG	282	
QY	121	GACGACCTTTTGTGAGATCCACGAAAGCACAAACGACCCACAGACACACAGCTACATT	180	
Db	283	GAAAGATTTCCGTGAGATTCATGCGCAAAACAAACGAGAGGACGAGATGACCATGTGTAACATT	342	
QY	181	TCCCGTGAGTTCCACCGCCGCTACCCGCTGCGGCTCCACAGTGTGACATATGCGCCCTCTCT	240	
Db	343	TCCCGTGAAATTTCACCGTGCCTACCCGCTGCGGCTTCCAAATGTGACCAATGCGCCCTCTCTCC	402	
QY	241	TGCTCCCGTGTGCGCGATGCGATGTGACTTGTGTGTGACCCCAAGATCCAGACTGGCGCTG	300	
Db	403	TGCTCTCTTGTCTGGGAGTGGATGTGACTTCTCTGTGCCCCAAGTCCAGTGTGGCTTG	462	
QY	301	GATGCCACCCACGCGCGAGCGACCATCCCGTGTGCGGGAGAGAGAGGCCCACTTGGCT	360	
Db	463	GATGTGGCCACAGCGAGAGAGGACCATTCCTCCGTGTCAACGGGAGAGAGAACCCAGTCTGGCA	522	
QY	361	CCCTGTCTCTGA	372	
Db	523	CCCTGTCTCTGA	534	

```

RESULT 4 316-253-110
US-10-316-253-110
; Sequence 110, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295

```


PRIOR FILING DATE: 2002-02-08
NUMBER OF SEQ ID NOS: 308
SOFTWARE: PatentIn version 3.1
SEQ ID NO 110
LENGTH: 1271
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
NAME/KEY: CDS
LOCATION: (159)..(749)
OTHER INFORMATION:
US-10-316-253-110

Query Match 70.7%; Score 263; DB 15; Length 1271;
Best Local Similarity 86.6%; Pred. No. 1.9e-67;
Matches 290; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 38 AGGTGATCCGACCGGGAACAAGTTCATCTTCTTCGATGTAGACATTTCTCCCGG 97
DB 415 AGGTGATCTGACCGGGAACAAGTTCATCTTCTTCGATGTAGACATTTCTCTCTG 474
QY 98 AGGACCTCACCCTGTAAGTGCAGACGACTTTGTGAGATCCAGAAAGACAAAGAGC 157
DB 475 AGGACCTCACCCTGTAAGTGCAGACGACTTTGTGAGATCCAGAAAGACAAAGAGC 534
QY 158 GCCAAGACGACGAGCTAATTTCCGTAAGTTCACCGCCGCTACCGCTGCGTCCA 217
DB 535 GGGAGAGTACGACGAGCTAATTTCCGTAAGTTCACCGCCGCTACCGCTGCGTCCA 594
QY 218 ACGTGACGACGAGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 277
DB 595 ATGTGACGACGAGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 654
QY 278 GCCCAGATCCAGACTGCTGATGACCAACGAGGAGGAGGAGGAGGAGGAGGAGG 337
DB 655 GCCCAGATCCAGACTGCTGATGACCAACGAGGAGGAGGAGGAGGAGGAGGAGG 714
QY 338 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372
DB 715 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 749

RESULT 5
US-10-029-386-12881

Sequence 12881, Application US/10029386
Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David K.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G

FILE REFERENCE: AEOMICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 12881

LENGTH: 573

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR21.3

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8

OTHER INFORMATION: NT HIT: AF026952.1, EVALU0 0.00e+00

OTHER INFORMATION: EST HUMAN HIT: BF726856.1, EVALU0 0.00e+00

OTHER INFORMATION: SWISSPROT HIT: P02489, EVALU0 1.00e-32

US-10-029-386-12881

Query Match 57.9%; Score 215.4; DB 15; Length 573;

Best Local Similarity 97.3%; Pred. No. 1.7e-53;
Matches 219; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 148 CACAAGAGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 207
DB 191 CACAAGAGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 250
QY 208 CTGCGCTCAAGTGCAGAGCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 267
DB 251 CTGCGCTCAAGTGCAGAGCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
QY 268 ACCTTCTGAGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 327
DB 311 ACCTTCTGAGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 370
QY 328 CCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372
DB 371 CCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 415

RESULT 6
US-10-029-386-26581

Sequence 26581, Application US/10029386
Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David K.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C

FILE REFERENCE: AEOMICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 26581

LENGTH: 211

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR21.3

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8

OTHER INFORMATION: SWISSPROT HIT: P02489, EVALU0 3.00e-36

OTHER INFORMATION: EST HUMAN HIT: BF726859.1, EVALU0 1.00e-113

OTHER INFORMATION: NT HIT: g114780619, EVALU0 1.00e-115

US-10-029-386-26581

Query Match 56.7%; Score 211; DB 15; Length 211;

Best Local Similarity 100.0%; Pred. No. 3e-52; Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 CAGAGACGACGAGCTACATTTCCGTAAGTTCACCGCGCTACCGCTGCGTCCAC 219
DB 1 CAGAGACGACGAGCTACATTTCCGTAAGTTCACCGCGCTACCGCTGCGTCCAC 60
QY 220 GTGAGCAAGTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 279
DB 61 GTGAGCAAGTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
QY 280 CCCAAGTCAAGCTGCGCTGATGACGACGACGACGACGACGACGACGACGACGAC 339
DB 121 CCCAAGTCAAGCTGCGCTGATGACGACGACGACGACGACGACGACGACGACGAC 180
QY 340 GAGAGAGGCGGACGCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
DB 181 GAGAGAGGCGGACGCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 211

RESULT 7
US-10-152-319A-1574

```
; Sequence 1574, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1574
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_012935
US-10-152-319A-1574

Query Match      34.8%; Score 129.6; DB 12; Length 528;
Best Local Similarity 64.2%; Pred. No. 2.8e-28;
Matches 217; Conservative 0; Mismatches 109; Indels 12; Gaps 1;
```

```
QY 13 ACCGTGCTGAGACTCCGGCATCTCTGAGGTTGGATCCGACCGGACAAAGTTGTCATCTTC 72
DB 175 AGCTGATTGACACTGGGCTCTCAGAGATGCGTATGAGAAAGACAGGTTCTCTGTGAAC 234
QY 73 CTCGATGTGAAGCACTTCTCCCGGAGAGACCTCAACCTGGAAGTSCAGAGACATTTGTG 132
DB 235 CTGACGTGAAGCACTTCTCTCCAGAGAACTCAAGTCAAGGTTCTGGAGACGTGATT 294
QY 133 GAGATCCAGGAAGACACAGAGCGCCAGAGACGACCAAGCTACATTTCCCGTGAAGTTC 192
DB 295 GAGGTGACACGGGACACAGAAAGCGCCAGAGACGAACATGGCTTCATCCAGGGAAGTTC 354
QY 193 CACCGCGCTACCGGCTGCGCTCCCAAGTGAACGATCGGCGCTCTCTGCGCCCTGCT 252
DB 355 CACAGGAAGTACCGGATCCAGCGCGAGCGTGAATCTCTCACCATTAATCTTCCCTGTCA 414
QY 253 GCCATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGATCGGCTGGATGCCACCCAC 312
DB 415 TCGATGAGAGTCTCTACGTGTGAATGACCAAGAAACAG-----GCCCTGCGC 462
QY 313 GCCGAGCGAGCCATCCCGTGTGCGGAGGAGAAAGCC 350
DB 463 CCGAGCGGACCATTCCTCATCAACCGGTGAAGAAAGCC 500
```

```
RESULT 8
US-09-917-800A-1419
; Sequence 1419, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1419
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 M55534
US-09-917-800A-1419
```

```
Query Match      34.8%; Score 129.6; DB 9; Length 1247;
Best Local Similarity 64.2%; Pred. No. 3.1e-28;
Matches 217; Conservative 0; Mismatches 109; Indels 12; Gaps 1;
```

```
QY 13 ACCGTGCTGAGACTCCGGCATCTCTGAGGTTGGATCCGACCGGACAAAGTTGTCATCTTC 72
DB 757 AGCTGATTGACACTGGGCTCTCAGAGATGCGTATGAGAAAGACAGGTTCTCTGTGAAC 816
QY 73 CTCGATGTGAAGCACTTCTCCCGGAGAGACCTCAACCTGGAAGTSCAGAGACATTTGTG 132
DB 817 CTGACGTGAAGCACTTCTCTCCAGAGAACTCAAGTCAAGGTTCTGGAGACGTGATT 876
QY 133 GAGATCCAGGAAGACACAGAGCGCCAGAGACGACCAAGCTACATTTCCCGTGAAGTTC 192
DB 877 GAGGTGACACGGGACACAGAAAGCGCCAGAGACGAACATGGCTTCATCCAGGGAAGTTC 936
QY 193 CACCGCGCTACCGGCTGCGCTCCCAAGTGAACGATCGGCGCTCTCTGCGCCCTGCT 252
DB 937 CACAGGAAGTACCGGATCCAGCGCGAGCGTGAATCTCTCACCATTAATCTTCCCTGTCA 996
QY 253 GCCATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGATCGGCTGGATGCCACCCAC 312
DB 997 TCGATGAGAGTCTCTACGTGTGAATGACCAAGAAACAG-----GCCCTGCGC 1044
QY 313 GCCGAGCGAGCCATCCCGTGTGCGGAGGAGAAAGCC 350
DB 1045 CCGAGCGGACCATTCCTCATCAACCGGTGAAGAAAGCC 1082
```

RESULT 9

```
US-09-954-456-514
; Sequence 514, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 514
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-514

Query Match          33.2%; Score 123.6; DB 9; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGTGTGACCTCCGCACTCTGAGGTTGATCCGCGGCAAGTTCATC 69
DB 197 CCCAGCTGTTGACACTGACCTCTGAGATGGCGCTGGAAGGACAGGTTCTGTC 256
QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGAGACCTCAACGTAAGTGACGAGCACTT 129
DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGACCTCAACGTAAGTGAGTGTG 316
QY 130 GTGAGATCCACGAAAGCAACAGAGCGCCAGACGACCAAGCTATTTCCCGTGAG 189
DB 317 ATGAGGTGATGAAAGCAAGTGAAGAGCGCAGAGTGAATGTTTCACTCCAGGAG 376
QY 190 TTCCACCGCGCTACCGCCCTGCGCTCCAACTGAGCAAGTCCGCTCTTGTCTCCTG 249
DB 377 TTCCACGAGAAATACCGAGATCCAGCTGATGTAGACCTCTCAACATTACTTACCTG 436
QY 250 TTGCGGATGACATGCTGACCTTCTGAGCCCAAGTCCAG 291
DB 437 TCATCTGATGGGCTCTCACTGTGATGACCAAGGAAACAG 478

RESULT 10
US-09-960-706-869
; Sequence 869, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William B.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplas
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
```

```
US-09-960-706-869
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 869
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 S45630
US-09-960-706-869

Query Match          33.2%; Score 123.6; DB 10; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGTGTGACCTCCGCACTCTGAGGTTGATCCGCGGCAAGTTCATC 69
DB 197 CCCAGCTGTTGACACTGACCTCTGAGATGGCGCTGGAAGGACAGGTTCTGTC 256
QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGAGACCTCAACGTAAGTGACGAGCACTT 129
DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGACCTCAACGTAAGTGAGTGTG 316
QY 130 GTGAGATCCACGAAAGCAACAGAGCGCCAGACGACCAAGCTATTTCCCGTGAG 189
DB 317 ATGAGGTGATGAAAGCAAGTGAAGAGCGCAGAGTGAATGTTTCACTCCAGGAG 376
QY 190 TTCCACCGCGCTACCGCCCTGCGCTCCAACTGAGCAAGTCCGCTCTTGTCTCCTG 249
DB 377 TTCCACGAGAAATACCGAGATCCAGCTGATGTAGACCTCTCAACATTACTTACCTG 436
QY 250 TTGCGGATGACATGCTGACCTTCTGAGCCCAAGTCCAG 291
DB 437 TCATCTGATGGGCTCTCACTGTGAAATGACCAAGGAAACAG 478

RESULT 11
US-09-873-319-566
; Sequence 566, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William B.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 566
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 S45630
US-09-873-319-566

Query Match          33.2%; Score 123.6; DB 10; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGTGTGACCTCCGCACTCTGAGGTTGATCCGCGGCAAGTTCATC 69
```

Db 197 CCAGCTGTTGACACTGACTCTCTCAGAGATGCGCTGAGAAAGACAGTTCTCTGTC 256
QY 70 TTCTCGATGTAAGCACTTCTCCCGGAGAGACTCAACCGTAAGTGACAGAGACTTT 129
Db 257 AACCTGATGTAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTGTGGAGATGTG 316
QY 130 GTGAGATCCAGGAAAGACACAGAGCGCCAGAGACGACCGCTACATTTCCCGTGAG 189
Db 317 ATTGAGGTGATGAGAAACATGAAGAGCGCCAGATGAATGTTTCATCTCCAGGGAG 376
QY 190 TTCCAGCGCGCTACCGCGCTGCGCTCCAGAGTGAACAGTGCGGCGCTCTTGTCCTG 249
Db 377 TTCCACAGGAAATACCGGATCCAGCTGATGAGACCTCTCCACCATTAATTCACTCG 436
QY 250 TCTGCCATGAGCAGTGCAGCTTCTGTGSCCCCAAGATCCAG 291
Db 437 TCATCTGATGGGGTCTCTCACTGTGAATGACCAAGAAACAG 478

RESULT 12

US-09-873-367C-764
; Sequence 764, Application US/09873367C
; Publication No. US20030165839A1

GENERAL INFORMATION:

APPLICANT: Young, Paul
APPLICANT: Soppet, Daniel
APPLICANT: Andrews, Gregory
APPLICANT: Augustus, Meena
APPLICANT: Ebner, Reinhard
APPLICANT: Carter, Kenneth
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
FILE OF INVENTION: Signature Gene Sets
FILE REFERENCE: 689290-64
CURRENT APPLICATION NUMBER: US/09/873,367C
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: U.S. 60/236,891
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/236,842
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/244,867
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: U.S. 60/245,084
PRIOR FILING DATE: 2000-11-01
NUMBER OF SEQ ID NOS: 1067
SOFTWARE: PatentIn version 3.0
SEQ ID NO 764
LENGTH: 691
TYPE: DNA
ORGANISM: Homo sapiens
US-09-873-367C-764

Query Match 33.2%; Score 123.6; DB 10; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;

Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGTCTGAGCTCCGGATCTCTGAGGTTGATCCGACCGGGAACAAGTTGCTATC 69
Db 197 CCAGCTGTTGACACTGACTCTCAGAGATGCGCTGAGAAAGACAGTTCTCTGTC 256
QY 70 TTCTCGATGTAAGCACTTCTCCCGGAGAGACTCAACCGTAAGTGACAGAGACTTT 129
Db 257 AACCTGATGTAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTGTGGAGATGTG 316
QY 130 GTGAGATCCAGGAAAGACACAGAGCGCCAGAGACGACCGCTACATTTCCCGTGAG 189
Db 317 ATTGAGGTGATGAGAAACATGAAGAGCGCCAGATGAATGTTTCATCTCCAGGGAG 376
QY 190 TTCCAGCGCGCTACCGCGCTGCGCTCCAGAGTGAACAGTGCGGCGCTCTTGTCCTG 249
Db 377 TTCCACAGGAAATACCGGATCCAGCTGATGAGACCTCTCCACCATTAATTCACTCG 436
QY 250 TCTGCCATGAGCAGTGCAGCTTCTGTGSCCCCAAGATCCAG 291

Db 437 TCATCTGATGGGGTCTCTCACTGTGAATGACCAAGAAACAG 478

RESULT 13

US-10-342-887-686
; Sequence 686, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:

APPLICANT: Dai, Hongyue
APPLICANT: He, Yudong
APPLICANT: Linsley, Peter S.
APPLICANT: Mao, Mao
APPLICANT: Roberts, Christopher J.
APPLICANT: Van 't Veer, Laura Johanna
APPLICANT: Van de Vijver, Marc J.
APPLICANT: Bernards, Rene

TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-188-999

CURRENT APPLICATION NUMBER: US/10/342,887
CURRENT FILING DATE: 2003-01-15

PRIOR APPLICATION NUMBER: 60/298,918
PRIOR FILING DATE: 2001-06-18

PRIOR APPLICATION NUMBER: 60/380,710
PRIOR FILING DATE: 2002-05-14

PRIOR APPLICATION NUMBER: 10/172,118
PRIOR FILING DATE: 2002-06-14

NUMBER OF SEQ ID NOS: 2699
SEQ ID NO 686

LENGTH: 691
TYPE: DNA

ORGANISM: Homo sapiens
US-10-342-887-686

Query Match 33.2%; Score 123.6; DB 13; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGTCTGAGCTCCGGATCTCTGAGGTTGATCCGACCGGGAACAAGTTGCTATC 69
Db 197 CCAGCTGTTGACACTGACTCTCAGAGATGCGCTGAGAAAGACAGTTCTCTGTC 256
QY 70 TTCTCGATGTAAGCACTTCTCCCGGAGAGACTCAACCGTAAGTGACAGAGACTTT 129
Db 257 AACCTGATGTAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTGTGGAGATGTG 316
QY 130 GTGAGATCCAGGAAAGACACAGAGCGCCAGAGACGACCGCTACATTTCCCGTGAG 189
Db 317 ATTGAGGTGATGAGAAACATGAAGAGCGCCAGATGAATGTTTCATCTCCAGGGAG 376
QY 190 TTCCAGCGCGCTACCGCGCTGCGCTCCAGAGTGAACAGTGCGGCGCTCTTGTCCTG 249
Db 377 TTCCACAGGAAATACCGGATCCAGCTGATGAGACCTCTCCACCATTAATTCACTCG 436
QY 250 TCTGCCATGAGCAGTGCAGCTTCTGTGSCCCCAAGATCCAG 291
Db 437 TCATCTGATGGGGTCTCTCACTGTGAATGACCAAGAAACAG 478

RESULT 14

US-10-172-118-686
; Sequence 686, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:

APPLICANT: Dai, Hongyue
APPLICANT: He, Yudong
APPLICANT: Linsley, Peter
APPLICANT: Mao, Mao
APPLICANT: Roberts, Chris

APPLICANT: Van 't Veer, Laura
APPLICANT: Van de Vijver, Marc
APPLICANT: Bernards, Rene
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-175-999

```
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 686
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_001885
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-686
```

```
Query Match      33.2%; Score 123.6; DB 13; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
```

```
QY 10 CGACCGTGTGAGCTCCGCGATCTTGAGTTGATCCGACCGGACAAGTTGTCATC 69
DB 197 CCAGCTGTTTGAAGTCACTCTCAGAGATGCGCTGAGAGACAGGTTCTCTGTC 256
QY 70 TTCCTGATGTGAAGCACTTCTCCCGAGAGACTCACCGTGAAGTGAAGAGACTT 129
DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGACTCAAAGTTAAGTGTGGAGATGTG 316
QY 130 GTGAGATCCAGGAAAGCAAAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 189
DB 317 ATTGAGGTGATGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 376
QY 190 TTCCACCGCGGCTACCGGCTGCGCTCCAGAGTGAAGTGAAGTGAAGTGAAG 249
DB 377 TTCCACAGGAAATACCGGATCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 436
QY 250 TCTGCCGATGAGCATGCTGACCTCTGAGGCCCAAGATCCAG 291
DB 437 TCATCTGATGGGTCTCTCACTGTGAATGACCAAGAAACAG 478
```

RESULT 15

```
US-10-133-937-61
; Sequence 61, Application US/10133937
; Publication No. US20030207278A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Melitzer, Paul
; TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
; TITLE OF INVENTION: DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
; FILE REFERENCE: 11613.56US01
; CURRENT APPLICATION NUMBER: US/10/133,937
; CURRENT FILING DATE: 2002-11-04
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-133-937-61
```

```
Query Match      33.2%; Score 123.6; DB 16; Length 691;
Best Local Similarity 64.9%; Pred. No. 1.7e-26;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
```

```
QY 10 CGACCGTGTGAGCTCCGCGATCTTGAGTTGATCCGACCGGACAAGTTGTCATC 69
DB 197 CCAGCTGTTTGAAGTCACTCTCAGAGATGCGCTGAGAGACAGGTTCTCTGTC 256
QY 70 TTCCTGATGTGAAGCACTTCTCCCGAGAGACTCACCGTGAAGTGAAGAGAGAGAG 129
DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGACTCAAAGTTAAGTGTGGAGATGTG 316
```

```
QY 130 GTGAGATCCAGGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
DB 317 ATTGAGGTGATGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 376
QY 190 TTCCACCGCGGCTACCGGCTGCGCTCCAGAGTGAAGTGAAGTGAAGTGAAG 249
DB 377 TTCCACAGGAAATACCGGATCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 436
QY 250 TCTGCCGATGAGCATGCTGACCTCTGAGGCCCAAGATCCAG 291
DB 437 TCATCTGATGGGTCTCTCACTGTGAATGACCAAGAAACAG 478
```

```
Search completed: September 27, 2004, 14:39:44
Job time : 3846 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2004, 12:02:53 ; Search time 333 Seconds

(without alignments)
4745.735 Million cell updates/sec

Title: US-10-657-740-2

Perfect score: 372
Sequence: 1 tccctctccgcacgcgtgct.....cctcggtccctcgtcctaa 372

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	315.2	84.7	531	7	ABX12062 Human alp
2	295.2	79.4	1056	9	ADBS2521 Primary r
3	129.6	34.8	528	7	ABT41872 Toxicity r
4	129.6	34.8	528	9	ADBS2642 Primary r
5	129.6	34.8	1247	6	ABK63512 Rat seque
6	123.6	33.5	537	9	ADBS75374 Human mut
7	123.6	33.2	537	2	ADBS75375 Human wil
8	123.6	33.2	691	9	AAX39668 Renal can
9	123.6	33.2	691	6	ABL65204 Lung can
10	123.6	33.2	691	6	ABL65204 Lung can
11	123.6	33.2	691	6	ABL65204 Lung can
12	123.6	33.2	691	6	ABL65204 Lung can
13	123.6	33.2	856	3	ABK64671 Human den
14	123.6	33.2	893	3	AAC03893 Human sec
15	123.6	33.2	911	3	AAC10863 Human sec
16	123.6	33.2	913	3	AAC10863 Human sec
17	123.6	33.2	927	3	AAC10863 Human sec
18	123.6	33.2	942	3	AAC10863 Human sec
19	123.6	33.2	1036	2	AAC03892 Human sec
20	123.6	33.1	548	6	ABZ35138 Human bre
21	117	31.5	449	7	ABX38978 Bovine ES
22	116.2	31.2	575	6	ABX39112 Bovine ES
23	116	31.2	575	6	ABX39112 Bovine ES

24	116	31.2	1771	3	AAC76681 Human ORF
25	115.6	31.1	599	6	ABV94670 Human pan
26	115.6	31.1	618	2	AAQ48718 Encodes p
27	115.6	31.1	865	6	ABO60780
28	115.6	31.1	1231	6	ABN97370 Human HSB
29	115.6	31.1	1231	9	ADD70998 Human hea
30	115.6	31.1	1380	3	AAA93441 GFP-HSP27
31	115.6	31.1	1380	6	ABX71560 DNA encod
32	115	30.9	695	3	AAC03894 Human sec
33	114.4	30.8	1488	2	AAZ42230 Human sec
34	110	29.6	500	8	ACH34922 Human nor
35	106.8	28.7	1310	7	ABT41710 Toxicity
36	106.8	28.7	1700	4	AAS56378 Human CDN
37	104.6	28.1	719	4	ABL07843 Drosophila
38	103.4	27.8	349	6	ABK67386 Target mo
39	102.8	27.6	450	8	ACH17737 Human adu
40	102.6	27.6	396	7	ABX47457 Bovine ES
41	101.2	27.2	503	6	ABL99507 Target ca
42	100.4	27.0	405	8	ACH16863 Human adu
43	98.6	26.5	376	4	AAL19387 Human bre
44	96.4	25.9	396	6	ABL63575 Breast ca
45	96.4	25.9	396	6	ABL63995 Breast ca

ALIGNMENTS

RESULT 1

ID ABX12062 standard, DNA, 531 BP.

AC ABX12062;

DT 16-MAY-2003 (first entry)

DE Human alpha BNAc crystallin chimera DNA.

XX Human; ds: gene: chimera; alpha BNAc crystallin; protein shelf life;
KW protein aggregation; accessible hydrophobic region increases; mutant;
KW larger size oligomer formation; interubunit interaction increase;
KW larger aggregate formation; larger porous oligomer formation;
KW increased ellipticity; less solvent accessible tryptophan;
KW increased chaperone-like activity; alpha A crystallin;
KW alpha B crystallin.

OS Homo sapiens.
XX Synthetic.

XX Key Location/Qualifiers

FT CDS 1..531

FT /tag= a "Alpha BNAc crystallin"

FT /partial

FT /note= "No stop codon given"

FT misc_feature 1..247

FT /tag= b

FT /note= "Derived from DNA sequence encoding N-terminal

FT misc_feature 248..531

FT /tag= c

FT /note= "Derived from DNA sequence encoding N-terminal

FT sequence of alpha A crystallin"

US2002177192-A1.

28-NOV-2002.

26-MAR-2002; 2002US-00105427.

28-MAR-2001; 2001US-0279223P.

(KUMA/) KUMAR L V S.

(RAOC/) RAO C M.

XX Kumar LVS, Rao CM;
XX WPI: 2003-298776/29.
DR P-PSDB; ABG76084.
XX
PT New chimera alpha BNAc nucleic acid, useful for preventing aggregation of
PT proteins and also for increasing shelf life of proteins of pharmaceutical
PT value.
XX
PS Claim 1, Fig 9; 17pp; English.
XX
CC The invention relates to a chimera alpha BNAc polynucleotide that encodes
CC a chimeric alpha BNAc polypeptide. The polypeptide is useful for
CC preventing protein aggregation. The polypeptide is also useful for
CC increasing the shelf life of proteins of pharmaceutical value. The
CC polypeptide shows an increase in accessible hydrophobic regions, forms
CC larger size oligomers, shows an increase in intermolecular interaction,
CC forms larger aggregates, forms larger porous oligomers and shows
CC increased ellipticity as compared to eye lens crystallins alpha A and
CC alpha B. The tryptophan residues in the polypeptide are less solvent
CC accessible as compared to those of eye lens crystallins alpha A and
CC B. The polypeptide shows extraordinarily high chaperone-like activity
CC ranging between 3 - 6 times that of the eye lens crystallins alpha A and
CC alpha B. The present sequence represents the chimera DNA that encodes
CC human alpha BNAc crystallin
XX
SQ Sequence 531 BP; 90 A; 198 C; 130 G; 113 T; 0 U; 0 Other;
XX
Query Match 84.7%; Score 315.2; DB 7; Length 531;
Best Local Similarity 92.2%; Pred. No. 2.2e-62;
Matches 332; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
XX
QY 10 CGCACCGGCTGGAGCTCCGCGATCTCTGAGGTCGACCGGACCAAGTTCGTCATC 69
DB 172 CCGAGCTGTTTGACACTGAGCTCTCAGAGATGCGCCCTGGAGAGACAGTTCTGTC 231
XX
QY 70 TTCTCGATGTAAGCACTTCTCCCGGAGACCTCACCGTAAAGTGAGAGACGACTTT 129
DB 232 AACCTGATGTGAAGCACTTCTCCCGGAGACCTCACCGTAAAGTGAGAGACGACTTT 291
XX
QY 130 GTGAGATCCACGGAAGACACACAGAGGCGCAAGACCAAGGTATATTTCCCGTAG 189
DB 292 GTGAGATCCACGGAAGACACACAGAGGCGCAAGACCAAGGTATATTTCCCGTAG 351
XX
QY 190 TTCCACCGCGCTACCGCTGCGCCGCAACGTCGAGCGGCTCTTGTGCTCCCTG 249
DB 352 TTCCACCGCGCTACCGCTGCGCCGCAACGTCGAGCGGCTCTTGTGCTCCCTG 411
XX
QY 250 TCTGCGGATGAGCATGCTGACCTTCTGTGAGCCCAAGATTCAGACTGCTTGATGCCACG 309
DB 412 TCTGCGGATGAGCATGCTGACCTTCTGTGAGCCCAAGATTCAGACTGCTTGATGCCACG 471
XX
QY 310 CACGCGGAGCGAGCATTCCTCGTGTGCGGAGGAGGAGCCCACTCGGCTCCCTGCTCC 369
DB 472 CACGCGGAGCGAGCATTCCTCGTGTGCGGAGGAGGAGCCCACTCGGCTCCCTGCTCC 531
XX
RESULT 2
ADBS2521
ID ADBS2521 standard; DNA, 1056 BP.
XX
AC ADBS2521;
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3063.
XX
KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KM toxicity marker; toxicity progression; drug screening;
KM primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.

XX WO2003065993-A2.
PN
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003W0-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442960P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
XX WPI: 2003-731472/69.
XX
XX Determining if a compound induces a toxic effect on a tissue or cell, for
XX identifying hepatotoxic compounds, comprises comparing a gene expression
XX profile of a tissue or cell sample to a database of Tox mean and non-Tox
XX mean values.
XX
XX Claim 44; SEQ ID NO 3063; 874bp; English.
XX
CC The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
SQ Sequence 1056 BP; 199 A; 333 C; 269 G; 255 T; 0 U; 0 Other;
XX
Query Match 79.4%; Score 295.2; DB 9; Length 1056;
Best Local Similarity 87.1%; Pred. No. 8.4e-58;
Matches 324; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
XX
QY 1 TCCCTCTCCGACCGGCTGGAGCTCCGGCATCTCTGAGGTTGATCCGACCGGAGCAAG 60
DB 163 TCTCTCTCCGACCGGCTGGAGCTCCGGCATCTCTGAGGTTGATCCGACCGGAGCAAG 222
XX
QY 61 TTGCGATCTTCTCGATGTAAGCACTTCTCCCGGAGACCTTACCGGTGAAGGTGAG 120
DB 223 TTGCGATCTTCTCGATGTAAGCACTTCTCTCTGAGACCTTACCGGTGAAGGTGAG 282
XX
QY 121 GACGACTTGTGAGATCCACGAAAGACACACAGGCGCCAGAGAGACCAACGCGCTACACT 180
DB 283 GAAAGTTTCGTGAGATCCATGCAACACACAGAGGCGAGGATGACCATGCTTACTT 342
XX


```

QY 181 TCCCGTAGTTCACCGCCGCTACCGCTGCGCTCCCAACGTGACCACTGGCCCTCTCT 240
DB 343 TCCCGTAGTTCACCGCTGCTACCGCTGCTCTTCCATGTGACCACTGGCCCTCTCC 402
QY 241 TGCTCCCTGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 403 TGCTCCCTGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462
QY 301 GATGCCACCGACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
DB 463 GATGCTGCGCCACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 522
QY 361 CCCCGCTGCTTA 372
DB 523 CCCCGCTGCTTA 534

RESULT 3
ABT41872
ID ABT41872 standard; DNA, 528 BP.
XX
AC ABT41872;
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 1574.
XX
KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.
XX
OS Rattus norvegicus.
XX
PN WO200295000-A2.
XX
PD 28-NOV-2002.
XX
PF 22-MAY-2002; 2002WO-US016173.
XX
PR 22-MAY-2001; 2001US-0292335P.
PR 13-JUN-2001; 2001US-0297523P.
PR 19-JUN-2001; 2001US-0298925P.
PR 10-JUL-2001; 2001US-0303807P.
PR 10-JUL-2001; 2001US-0303808P.
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 27-SEP-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 06-DEC-2001; 2001US-0336144P.
PR 19-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.
PR 21-FEB-2002; 2002US-0357844P.
PR 15-MAR-2002; 2002US-0364134P.
PR 08-APR-2002; 2002US-0370144P.
PR 08-APR-2002; 2002US-0370206P.
PR 08-APR-2002; 2002US-0370247P.
PR 17-APR-2002; 2002US-0372794P.
PR 21-APR-2002; 2002US-0371679P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX
XX WPI; 2003-148464/14.
XX
PT Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
PS Example 4; Page: 446pp; English.

```

```

XX The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene or
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the renal
CC progression of a toxic effect of a compound, predicting the renal
CC toxicity of a compound, or identifying toxicity markers in tissues or
CC cells exposed to known renal toxins. The genes are useful as toxicity
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 528 BP, 114 A; 170 C; 128 G; 116 T; 0 U; 0 Other;
XX
Query Match 34.8%; Score 129.6; DB 7; Length 528;
Best Local Similarity 64.2%; Pred. No. 3.3e-20;
Matches 217; Conservative 0; Mismatches 109; Indels 12; Gaps 1;
QY 13 ACCGTGCTGACTCCGGCATCTGTGAGTTGATCCGACCGGACAACTGTCATCTTC 72
DB 175 AGCTGATGTACACTGGGCTCTCAGAGATCGTATGAGAAAGCAAGTTCTCTGTGAC 234
QY 73 CTGATGTGAGACATTTTCTCCCGAGAGACTCAACCTGAAAGTGACAGACATTTGTG 132
DB 235 CTGACGTGAAAGCACTTCTTCCAGAGAACTCAAAATCAAGTTCTGGAGACGTATT 294
QY 133 GAGATCCAGGAAAGCAACAGACGACGACGACGACGACGACGACGACGACGACGACGAC 192
DB 295 GAGGTGACGCGCAAGACAGACGACGACGACGACGACGACGACGACGACGACGACGAC 354
QY 193 CACCGCGCTACCGCCTGCGCTCCAACTGAGACGACGACGACGACGACGACGACGACGAC 252
DB 355 CACAGAAAGTACCGGATCCAGCGCGAGCTGATCTCTCACTTCTTCCCTGTCA 414
QY 253 GCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 312
DB 415 TCGGATGAGTCTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462
QY 313 GCCGAGGAGGACCATCCCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 350
DB 463 CCGAGGACGACCATTTCCCATCACTCCGTAAGAGAGAGCC 500

RESULT 4
ADBS2642
ID ADBS2642 standard; DNA, 528 BP.
XX
AC ADBS2642;
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3184.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
PN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003WO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.

```

PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394230P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
XX WPI, 2003-731472/69.
XX
XX
XX Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
XX
XX Claim 44; SEQ ID NO 3184; 874pp; English.
XX
XX
XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
XX
XX Sequence 528 BP; 114 A; 170 C; 128 G; 116 T; 0 U; 0 Other;
SQ
Query Match 34.8%; Score 129.6; DB 9; Length 528;
Best Local Similarity 64.2%; Pred. No. 3.3e-20;
Matches 217; Conservative 0; Mismatches 109; Indels 12; Gaps 1
QY 13 ACCGTGCTGACTCCGGCATCTCTGTGAGTTGCATCCGACCGGACAAAGTTGTCATCTTC 72
Db 175 AGCTGATGTAACACACGCGGCTCTCAGAGAACCGATGAGAAAGACAGGTTCTGTGAAC 234
QY 73 CTCGATGTGAAGCACTTTCCTCCCGGAGAGACTCACCGTAGAGTGCAGCACTTGTG 132
Db 235 CTGACGCTGAGACACTTCTCTCTCCAGAGAACTCAAGTCAAGGTTCTGGGAGACGTGATT 294
QY 133 GAGATCCACGGAAAGCACAAAGAGCGCCAGAGACGACCAAGCTACATTTCCGTGAGTTG 192
Db 295 GAGGTGACACGGCAACACAGAAAGCGCCACAGAGCAACATGAGTTCATCTCCAGAGGATT 354
QY 123 CACGCGCCGCTACCGCTGCGCTGCACACGTGAGACCAATCGGCGCCTCTCTGTCGCTGTC 252
Db 355 CACAGAGAGTACCGGATTCGACGCGACGAGTGATCTCTCAACCAATTACTTTCTCCCTGTCA 414
QY 253 GCCGATGGCATCTGACCTTCTGTGGCCCCCAAGATCCAGATCGGCTGATGACCCAC 312
Db 415 TCGATGTGAGTCTCTCACTGTGATGAGCAACAGAAACAG-----GCCTCGGC 462
QY 313 GCCGAGCGGACCATCCCGCTGTGCGGAGAGAGAGACGCC 350

Db 463 CCTGAGCGCACCATTCCTATCACCCTGTGAAGAGCC 500

RESULT 5
ABK63512
ID ABR63512 standard; cDNA, 1247 BP.
XX
XX ABR63512;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #1419.
XX
KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
XX differential expression; centrilobular necrosis; steatosis.
XX
CS Rattus norvegicus.
XX
XX W0200210453-A2.
XX
XX 07-FEB-2002.
XX
PD 30-JUL-2001; 2001WO-US023872.
XX
PF 31-JUL-2000; 2000US-0222040P.
XX
PR 02-NOV-2000; 2000US-0244880P.
XX
PR 11-MAY-2001; 2001US-0230025P.
XX
PR 15-MAY-2001; 2001US-0230645P.
XX
PR 22-MAY-2001; 2001US-0232336P.
XX
PR 06-JUN-2001; 2001US-0235796P.
XX
PR 13-JUN-2001; 2001US-0237457P.
XX
PR 19-JUN-2001; 2001US-0239884P.
XX
PR 09-JUL-2001; 2001US-0303459P.
XX
XX (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX
XX WPI; 2002-241625/29.
XX
DT
XX
XX Predicting toxic effects of compounds or the progression of these toxic
XX effects by determining the changes in gene expression in tissues or cells
XX exposed to the toxin and comparing these to gene expression in unexposed
XX tissues or cells.
XX
XX
XX Claim 1; SEQ ID NO 1419; 239pp; English.
XX
XX The invention relates to methods for predicting toxic effects of
XX compounds or the progression of these toxic effects by determining the
XX global changes in gene expression in tissues or cells exposed to the
XX toxin and comparing these to gene expression in unexposed tissues or
XX cells. Also included are methods of predicting at least one toxic effect
XX of a compound or progression of a toxic effect, preferably the
XX hepatotoxicity of a compound, comprising detecting the level of
XX expression in a tissue or cell sample exposed to the compound of two or
XX more genes listed in the specification, where differential expression of
XX the genes is indicative of at least one toxic effect or progression. The
XX method can also be used to identify an agent which modulates the toxic
XX response and predict cellular pathways that a compound modulates in a
XX cell. The methods utilise a set of at least two probes (on a solid
XX support in kit form), where each of the probes comprises a sequence that
XX specifically hybridises to a gene listed in the specification, a computer
XX system comprising a database containing information identifying the
XX expression level in a tissue or cell sample exposed to a hepatotoxin of a
XX set of genes comprising at least two genes listed in the specification,
XX and a user interface to view the information used to present information,
XX identifying the expression level in a tissue or cell of at least one gene
XX listed in the specification. The method is useful for elucidating global
XX changes in gene expression and for identifying toxicity markers in
XX tissues or cell exposed to a known toxin. The genes may be used as
XX toxicity markers in drug screening and toxicity assays. The genes and
XX gene expression information may be used as diagnostic markers for the

CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
XX Sequence 1247 BP; 294 A; 360 C; 306 G; 287 T; 0 U; 0 Other;

Query Match 34.8%; Score 129.6; DB 6; Length 1247;
Best Local Similarity 64.2%; Pred. No. 3.7e-20;
Matches 217; Conservative 0; Mismatches 109; Indels 12; Gaps 1;

QY 13 ACCGTGCTGGACCTCCGAGATCTTGAGTTGATCCAGCCGGACAAAGTTGTCATCTTC 72
DB 757 AGCTGATGTACACTGGGCTCTCAGAGATGGTATGAGAGAGACAGGTTCTCTGTGAAC 816
QY 73 CTCGATGTGAAGCACTTCTCCCGAGACCTCAGCGTGAAGTGGAGAGACGATTTGTG 132
DB 817 CTGACGTGAAGCACTTCTCTCCAGAGAACTCAAAGTCAAGTTCTGGAGAGCTGATT 876
QY 133 GAGATCCAGGAAAGCAAGACGAGCGGACAGACGACGCTACATTTCCGTTGAGTTTC 192
DB 877 GAGGTGACGCGCAAGACAGAAAGCGCCAGAGACGACATGCTTCATCTCAGGAGATTTC 936
QY 193 GACCGCGGTAACGCGCTGCGCTCCAGCTGAGACCAATCGCCCTCTTGTCTCTGCTCT 252
DB 937 CAGAGAAATACCGGATCCAGCGAGCTGATCTCTCCACATTACTTCTCTCTGTCGA 996
QY 253 GCCGATGCGATGCTGACCTTCTGTGCGCCCAAGATCCAGACTGGCCGTGATGCCACCCAC 312
DB 997 TCGATGTGACTCTCTCACTGTGATGACCAAGAAACAG-----GCCCTGTGC 1044
QY 313 GCCGAGCGACCATCCCGTGTGCGGAGAGAAAGCC 350
DB 1045 CTTGAGGCGACCATTCCTCATCACCCTGAGAGAAAGGCC 1082

RESULT 6
ADE75374
ID ADE75374 standard; DNA; 537 BP.

XX ADE75374;

DT 29-JAN-2004 (first entry)

DE Human mutant alphab-crystallin fragment-encoding DNA.

XX
XX Hybrid protein chaperone; protein stabilisation; heat shock protein;
KM SHSP family; protein aggregation inhibition; cell death inhibition;
KM genome stability pathway inhibition; protein denaturation identification;
KM protein conformation related disease; cardiomyopathy; cataract;
KM neurodegenerative disease; cataract; ophthalmological; neuroprotective;
KM gene therapy; alphab-crystallin; wild-type; human; cassette mutagenesis;
mutant; ds.
XX

OS Synthetic.
OS Homo sapiens.

XX Key Location/Qualifiers

FT misc_feature 493..498
FT /tag= a
FT /note= "Unique Avar site for cassette mutagenesis"
FT mutation replace(497,T)
FT /tag= b
FT /note= "Mutagenesis to G from T at this site generates a
XX unique Avar site between bases 493-498"

PN MO2003091266-A2.

PD 06-NOV-2003.

PF 23-APR-2003; 2003WO-GB001721.

PR 23-APR-2002; 2002GB-00009334.

PA (UYDU-) UNIV DUNDEE.

XX Quinlan R;

XX WPI; 2003-865571/80.

DR New hybrid protein chaperone (e.g. heat shock protein) useful for
PT stabilizing proteins and/or protein activities, or as an agent to prevent
PT protein aggregation, or for treating diseases involving altered protein
PT conformations.

XX Disclosure; Fig 12; 45pp; English.

XX The invention relates to a hybrid protein chaperone for stabilising
CC proteins and/or protein activities. Protein chaperones (also known as
CC heat shock proteins) are divided into 4 families on the basis of their
CC primary sequence and chaperone properties: HSP90, HSP70, HSP60 and SHSP
CC (small heat shock protein). The invention is based upon the finding that
CC among the SHSP family, which have a general structure of a central domain
CC (called the alpha-crystallin domain) flanked by N and C-terminal regions,
CC replacement of one or more regions of an SHSP with the corresponding
CC region from a second SHSP can improve the activity compared to native
CC SHSPs. In a particular embodiment of the invention, the hybrid chaperone
CC is a hybrid SHSP designated alphab-HSP27 comprising the N-terminus and
CC central portion of alphab-crystallin and the C-terminal tail of HSP27.
CC However, the hybrid protein chaperones of the invention can comprise
CC regions from HSP90, HSP70 and HSP60 families as well as from the SHSP
CC family. The invention also encompasses methods for stabilising proteins
CC such as enzymes, therapeutic proteins, diagnostic proteins, reporter
CC proteins or antibodies, their fragments or conjugates in an aqueous
CC solution using hybrid protein chaperones; stabilised protein formulation
CC comprising at least one protein associated with the above hybrid protein
CC chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
CC and the coexpression of a recombinant protein of interest and a hybrid
CC protein chaperone. The hybrid protein chaperones are useful as agents to
CC prevent protein aggregation, as inhibitors of cell death and genome
CC stability pathways, for identification of proteins that are in the
CC process of unfolding, for the treatment of diseases involving altered
CC protein conformations (e.g., cardiomyopathies, cataracts or
CC neurodegenerative disease), or for the manufacture of a medicament for
CC the treatment of such diseases. The present sequence represents a mutant
CC human alphab-crystallin DNA sequence in which the G at position 497 was
CC mutated from T (see ADE75375) to generate a unique Avar restriction site
CC which was used for subsequent cassette mutagenesis (along with a vector-
CC based SacI site) to introduce heterologous C-terminal sequences onto
CC alphab-crystallin.
XX

SQ Sequence 537 BP; 118 A; 171 C; 125 G; 123 T; 0 U; 0 Other;

Query Match 33.5%; Score 124.6; DB 9; Length 537;
Best Local Similarity 63.0%; Pred. No. 4.6e-19;
Matches 215; Conservative 0; Mismatches 114; Indels 12; Gaps 1;

QY 10 CGACCGTGTCTGAGCTCCGAGATCTCTGAGGTTGATCCAGCCGGACAAAGTTCGTCATC 69
DB 180 CCAGCTGTTTGAACACTGTGACTCTCAGAGATCCGCTGGAAGAGACAGGTTCTCTGTC 239

QY 70 TTCTCTGATGTGAAGCACTTCTCCCGAGAGACTTCACCGTGAAGTGCAGACGACTTT 129

DB 240 AACCTGATGTGAAGCACTTCTCCAGAGAACTCAAAGTTAAGTGTGGAGATGTG 299

QY 130 GTGAGATCCAGGAAAGCAAGACGCGCCAGAGAGACCAACGCGCTACATTTCCCGTGA 189

DB 300 ATTGAGGTCAATGAAACATGAGAGCGCCAGATGAACATGAGTTTCACTCCAGGAG 359

QY 190 TTCCACCGCGCTACCGCTGCGCTGCTCAACGTGACCAAGTGGCCCTCTTGTCTCCGTG 249

DB 360 TTCCACAGGAAATACCGGATCCAGCTGATGTAGACCTCTACCACTTATCTTATCCCTG 419

QY 250 TCTGCCGATGAGCATGTGACTTCTGTGCGCCCAAGATCCAGACTGGCTGATGCCACC 309

Db 420 TCATCTGATGGGGTCTCCTCACTGTGATGACCAAGAAAGAG-----GTCTCT 467

QY 310 CACCGCGAGGAGCCATCCCGTGTCCGGAGAGAGAGCC 350
 |||||
 Db 468 GGCCCTGAGCGCACCATTCCTCCATCACCCTGGAGAGAGAGCC 508

RESULT 7
 ADE75375
 ID ADE75375 standard; DNA; 537 BP.
 AC ADE75375;
 AC
 XX 29-JAN-2004 (first entry)
 XX
 DE Human wild-type alphaB-crystallin fragment-encoding DNA.

XX Hybrid protein chaperone; protein stabilisation; heat shock protein;
 KW SHP family; protein aggregation inhibition; cell death inhibition;
 KW genome stability pathway inhibition; protein denaturation identification;
 KW protein conformation related disease; cardiomyopathy; cataract;
 KW neurodegenerative disease; cardiac; ophthalmological; neuroprotective;
 KW gene therapy; alphaB-crystallin; wild-type; human; cassette mutagenesis;
 KW ds.

XX Homo sapiens.
 OS
 XX
 PH Key location/Qualifiers
 FT mutation replace(497,G)
 FT /*tag= a
 FT /note= "Mutagenesis of T to G at this site generates a
 unique Aval site between bases 493-498"

XX W02003091266-A2.
 EN
 XX
 PD 06-NOV-2003.
 XX
 PF 23-APR-2003; 2003WO-GB001721.
 XX
 PR 23-APR-2002; 2002GB-00009334.
 XX
 XX (UYDU-) UNIV DUNDIE.
 PA
 PI Quinlan R;
 DR WPI; 2003-865571/80.
 XX
 XX
 PT New hybrid protein chaperone (e.g. heat shock protein) useful for
 PT stabilizing proteins and/or protein activities, or as an agent to prevent
 PT protein aggregation, or for treating diseases involving altered protein
 PT conformations.

XX PS Disclosure; Fig 12; 45pp; English.

XX The invention relates to a hybrid protein chaperone for stabilising
 CC proteins and/or protein activities. Protein chaperones (also known as
 CC heat shock proteins) are divided into 4 families on the basis of their
 CC primary sequence and chaperone properties: HSP90, HSP70, HSP60 and SHSP
 CC (small heat shock protein). The invention is based upon the finding that
 CC among the SHSP family, which have a general structure of a central domain
 CC (called the alpha-crystallin domain) flanked by N and C-terminal regions,
 CC replacement of one or more regions of an SHSP with the corresponding
 CC region from a second SHSP can improve the activity compared to native
 CC SHSPs. In a particular embodiment of the invention, the hybrid chaperone
 CC is a hybrid SHSP designated alphaB-HSP27 comprising the N-terminus and
 CC central portion of alphaB-crystallin and the C-terminal tail of HSP27.
 CC However, the hybrid protein chaperones of the invention can comprise
 CC regions from HSP90, HSP70 and HSP60 families as well as from the SHSP
 CC family. The invention also encompasses methods for stabilising proteins
 CC such as enzymes, therapeutic proteins, diagnostic proteins, reporter
 CC proteins or antibodies, their fragments or conjugates in an aqueous
 CC solution using hybrid protein chaperones; stabilised protein formulation
 CC comprising at least one protein associated with the above hybrid protein

CC chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
 CC and the coexpression of a recombinant protein of interest and a hybrid
 CC protein chaperone. The hybrid protein chaperones are useful as agents to
 CC prevent protein aggregation, as inhibitors of cell death and genome
 CC stability pathways, for identification of proteins that are in the
 CC process of unfolding, for the treatment of diseases involving altered
 CC protein conformations (e.g., cardiomyopathies, cataracts or
 CC neurodegenerative disease), or for the manufacture of a medicament for
 CC the treatment of such diseases. The present sequence represents a wild-
 CC type human alphaB-crystallin DNA sequence. The T at position 497 was
 CC mutated to G (see ADE75374) to generate a unique Aval restriction site
 CC which was used for subsequent cassette mutagenesis (along with a vector-
 CC based SacI site) to introduce heterologous C-terminal sequences onto
 CC alphaB-crystallin.
 CC
 XX
 SQ Sequence 537 BP; 118 A; 171 C; 124 G; 124 T; 0 U; 0 Other;

Query Match 33.2%; Score 123.6; DB 9; Length 537;
 Best Local Similarity 64.9%; Pred. No. 7.7e-19;
 Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGCACCGTGTGCACTCCGCACTCTGAGGTTCATCCGAGCAAGTTCATC 69
 |||||
 Db 180 CCAGCTGTTTGAACAATGACATCTCAGAGATGCGCTGAGAGAGACAGGTCTCTGTC 239

QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGGACTCACCGTGAAGTGCAGAGACTT 129
 |||||
 Db 240 AACCTGATGTGAAGCACTTCTCCCGAGGACTCAAAGTGAAGTGCAGAGACT 299

QY 130 GTGAGATGTCAAGCAAGCAACAGAGCGCCAGAGACCAAGGCTCATTTCCCTGAG 189
 |||||
 Db 300 ATTGAGTGTCAAGCAAGCAACAGAGCGCCAGAGACCAAGGCTCATTTCCCTGAG 359

QY 190 TTCCACCGCGCTACCGCCTGCGCTCCACGCTGAGCACTGCGCTCTTCTCTCCTG 249
 |||||
 Db 360 TTCCACAGGAAATATCCGAGATCCAGCTGATGTAGACCTTCACCATTAATTCTCCTG 419

QY 250 TCTGCCGATGCACTGCACTTCTGTGGCCCCCAAGATCCAG 291
 |||||
 Db 420 TCATCTGATGGGGTCTCCTCACTGTGATGACCAAGAAAGAG 461

RESULT 8
 AAX39668
 ID AAX39668 standard; DNA; 691 BP.
 XX
 AC AAX39668;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Renal cancer associated gene.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

XX Homo sapiens.
 OS
 XX
 PN W09904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US014679.
 XX
 PR 17-JUL-1997; 97US-00896164.
 PR 10-OCT-1997; 97US-0061599P.
 PR 10-OCT-1997; 97US-0061765P.
 PR 10-OCT-1997; 97US-00948705.
 PR 11-OCT-1997; 97GB-00021687.
 XX 22-JUN-1998; 98US-00102332.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.

CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour
 XX

Sequence 691 BP, 168 A, 208 C, 148 G, 167 T, 0 U, 0 Other;
 33.2%; Score 123.6; DB 6; Length 691;
 Best Local Similarity 64.9%; Pred. No. 8e-19;
 Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Query Match
 10 CGACCGTCTGGACCTCCGATCTTGAGTTGATCCGCGGACAGTTGTCATC 69
 197 CCGAGCTGTTGACCTGACCTTCAGAGTGGCCTGGAGAGAGAGTTCTGTGTC 256

Db
 197 TTCTCGATGGAAGCACTTCTCCCGAGAGCCTCAACGTAAGTGCAGAGCACTTT 129
 257 AACCTGATGGAAGCACTTCTCCCGAGAGCACTCAAGTTAAGTGTGGAGATGTG 316

Qy
 130 GTGAGATCCACGGAAGACACAGAGCGCCAGAGACCAACGCGCTACATTTCCCGTAG 189
 317 ATTGAGGTGATGAGAAACATGAAGAGCGCCAGATGAACTGTTTCACTCCAGGGAG 376

Db
 190 TTCCACCGCGCTACCGCGCTGCGCTCCAGGTGAGACCAAGTGGCCCTCTTGTGCTG 249
 377 TTCCACGAGAAATACCGGATCCAGCTGATGATGACCTCTCCACCATTTACTTCCTG 436

Qy
 250 TCTGCGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 291
 437 TCATCTGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 478

Db

RESULT 10
 ABL62427
 ID ABL62427 standard; DNA; 691 BP.
 XX ABL62427;
 AC
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Colon adenocarcinoma related gene sequence SEQ ID NO:764.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytosolic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 EN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 XX 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 28-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 XX (AVAL-) AVALON PHARM.
 XX
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 XX
 XX
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 XX
 PS Claim 1; SEQ ID NO 764; 44pp; English.
 PS
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytosolic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell,
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour
 CC
 XX
 SQ Sequence 691 BP, 168 A, 208 C, 148 G, 167 T, 0 U, 0 Other;
 33.2%; Score 123.6; DB 6; Length 691;
 Best Local Similarity 64.9%; Pred. No. 8e-19;
 Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Query Match
 10 CGACCGTCTGGACCTCCGATCTTGAGTTGATCCGCGGACAGTTGTCATC 69
 197 CCGAGCTGTTGACCTGACCTTCAGAGTGGCCTGGAGAGAGAGTTCTGTGTC 256

Db
 197 TTCTCGATGGAAGCACTTCTCCCGAGAGCCTCAACGTAAGTGCAGAGCACTTT 129
 257 AACCTGATGGAAGCACTTCTCCCGAGAGCACTCAAGTTAAGTGTGGAGATGTG 316

Qy
 130 GTGAGATCCACGGAAGACACAGAGCGCCAGAGACCAACGCGCTACATTTCCCGTAG 189
 317 ATTGAGGTGATGAGAAACATGAAGAGCGCCAGATGAACTGTTTCACTCCAGGGAG 376

QY 190 TTCACCGCGCTACCGGCTGCGTCCAGCTGAGACCAAGTGGCCCTCTTGTCTCCCTG 249
DB 377 TTTCACAGGAAATACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCAATCCCTG 436
QY 250 TCTGCCGATGCGATGCTGATCTTCTGTGGCCCAAGATCCAG 291
DB 437 TCATCTGATGGGGTCTCCTCAGTGAATGACCAAGAAACAG 478

RESULT 11
AB110966
ID AB110966 standard; cDNA, 691 BP.

AC AB110966;

DT 04-DEC-2002 (first entry)

DE Human breast cancer associated coding sequence SEQ ID NO: 1100.

KW Human; breast specific gene; breast cancer; differential expression;

KM cytosolic; gene therapy; gene; ss.

OS Homo sapiens.

PN WO200259271-A2.

PD 01-AUG-2002.

PF 25-JAN-2002; 2002WO-US002176.

PR 25-JAN-2001; 2001US-0263757P.

PR 25-APR-2001; 2001US-0286090P.

PR 23-MAY-2001; 2001US-0292517P.

PA (GENE-) GENE LOGIC INC.

PI Orr MS, Nation M, Diggans JC, Zeng W;

DR WPI; 2002-674803/72.

PT Diagnosing breast cancer in a patient comprises detecting the level of

PT gene expression in cell or tissue samples, where a differential gene

PT expression is indicative of breast cancer.

PS Claim 1; SEQ ID NO 1100; 260pp + Sequence Listing; English.

XX The present invention relates to methods of diagnosing breast cancer in a

CC patient, which comprise detecting the level of expression in a tissue

CC sample of two or more genes selected from those shown in AB10967-

CC AB11112, where a differential expression of the genes indicates breast

CC cancer. The methods are useful in diagnosing, treating, detecting the

CC progression, and in monitoring treatment of breast cancer in patients.

CC The methods are also useful as a screening tool for agents that modulate

CC the onset or progression of breast cancer. The breast cancer genes may be

CC used as diagnostic markers for the prediction or identification of the

CC malignant state of breast tissue, for confirming the type and progression

CC of cancer, and for drug screening and assays. The present sequence is a

CC coding sequence of the invention. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub.published_pct_sequences

XX Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

SO Query Match 33.2%; Score 123.6; DB 6; Length 691;

Best Local Similarity 64.9%; Pred. No. 8e-19; Mismatches 99; Indels 0; Gaps 0;

Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGACCGGCTGATCCGCGATCTCTGAGTTGATCCGACCGGCAAGTGTCTGATC 69

DB 197 CCGAGCTGTTGACACTGATCTCAAGATGCGCTCGGAGAGACAGTTCCTGTC 256

QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGAGACCTCACCCTGTAAGTGCAGAGACGACTTT 129

DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGTGTGGAGATGTG 316
QY 130 GTGAGATTCACGGAAGACCAACAGAGCGCCAGGACCAAGCAAGTATTCCTGAG 189
DB 317 ATTGAGTGTGATGGAAGCAATGAAAGCGCCAGATGAACATGTTTCATCTCAAGGAG 376
QY 190 TTCACCGCGCTACCGGCTGCGTCCAGCTGAGACCAAGTGGCCCTCTTGTCTCCCTG 249
DB 377 TTTCACAGGAAATACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCAATCCCTG 436
QY 250 TCTGCCGATGCGATGCTGATCTTCTGTGGCCCAAGATCCAG 291
DB 437 TCATCTGATGGGGTCTCCTCAGTGAATGACCAAGAAACAG 478

RESULT 12

ABK64671
ID ABK64671 standard; DNA, 691 BP.

AC ABK64671;

DT 18-JUN-2002 (first entry)

DE Human benign prostatic hyperplasia gene #566.

KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.

OS Homo sapiens.

PN WO200212440-A2.

PD 14-FEB-2002.

PF 07-AUG-2001; 2001WO-US024708.

PR 07-AUG-2000; 2000US-0223323P.

PR 05-JUN-2001; 2001US-00873319.

PA (GENE-) GENE LOGIC INC.

PI (NLSB) JAPAN TOBACCO INC.

PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;

DR WPI; 2002-257476/30.

PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by

PT detecting expression levels of one or more genes in prostate cells from

PT patient that are differentially regulated compared to normal prostate

PT cells.

PS Disclosure; Page 307-308; 444p; English.

XX The invention relates to a method of diagnosing (I) the onset or

CC progression of benign prostatic hyperplasia (BPH), or screening (II) for

CC or identifying an agent that modulates the onset or progression of BPH.

CC The method is based on changes in gene expression in BPH tissue isolated

CC from patients exhibiting different clinical states of prostate

CC hyperplasia as compared to normal prostate tissue. (I) comprises

CC detecting the expression levels of one or more genes in prostate cells

CC from the subject that are differentially regulated compared to normal

CC prostate cells. (II) comprises preparing a first gene expression profile

CC of BPH cells or BPH-like cell population, exposing the cells to the

CC agent, preparing a second gene expression profile of the agent exposed

CC cells, and comparing the first and second gene expression profiles. (I)

CC is useful for diagnosing the onset or progression of BPH. (II) is useful

CC for identifying an agent that modulates the onset or progression of BPH.

CC The methods are useful to present information identifying the expression

CC level in a tissue or cells, by comparing the expression level of genes

CC given in the specification in the database, and displaying the expression levels

CC of expression of gene in the tissue or cell sample compared to the

CC expression level in BPH. Agents using (II) are useful for treating BPH or

CC prostate cancer. ABK4106-ABK64860 represent human benign prostatic
CC hyperplasia gene sequences of the invention
XX
SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;
Query Match 33.2%; Score 123.6; DB 6; Length 691;
Best Local Similarity 64.9%; Pred. No. 8e-19;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 10 CGCACCGGCTGAGTCCGCGCATCTCTGAGCTTCGATCCGACCGGAGCAAGTTGCTCATC 69
DB 197 CCCAGCTGGTTGACACTGAGACTCTCAGAGATGCCCTGGAGAAAGACAGTTCTCTGTG 256
QY 70 TTCCTCGATGTGAAGCACTTCTCCCGAGAGACCTCACCCTGAGAGTGCAGACACTTT 129
DB 257 AACCTGATGTGAAGCACTTCTCCCGAGAGACCTCACCCTGAGAGTGCAGACACTTT 316
QY 130 GTGAGATCCACGGAAGACAGAGGCGCGAGGACGACGCTACATTTCCCGTAG 189
DB 317 ATTGAGGTGATGGAAGAAACATGAGAGCGCCAGATGACATGTTTCACTCCAGGAG 376
QY 190 TTCCACCGCCGCTACCGGCTGCGCTGCAACGTGACCAAGTCCCTCTCTTGTCTCCCTG 249
DB 377 TTCACAGGAAGAAATACCGGATCCGAGCTGATGAGACCTCTCACCATTACTTCACTCCTG 436
QY 250 TCTGCCGATGCGATGCTGACCTTCTGTGCGCCCAAGATCCAG 291
DB 437 TCATCTGATGAGGAGTCCCTCACTGTGATGAGACCAAGAAACAG 478

RESULT 13

AAC03893
ID AAC03893 standard; cDNA; 856 BP.

XX AAC03893;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 3891.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

OS

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR P-PSDB; AAG03887.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 3891; 71bp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX mRNAs encoding secreted proteins. An ORF has been identified within the

XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

XX derived from 30 different tissues. EST sequences usually correspond

XX mainly to the 3' untranslated region (UTR) of the mRNA because they are

XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

XX well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX
SQ Sequence 856 BP; 201 A; 255 C; 196 G; 197 T; 0 U; 7 Other;

Query Match 33.2%; Score 123.6; DB 3; Length 856;

Best Local Similarity 64.9%; Pred. No. 8.2e-19;

Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 CGCACCGGCTGAGTCCGCGCATCTCTGAGCTTCGATCCGACCGGAGCAAGTTGCTCATC 69
DB 400 CCCAGCTGGTTGACACTGAGACTCTCAGAGATGCCCTGGAGAAAGACAGTTCTCTGTG 459
QY 70 TTCCTCGATGTGAAGCACTTCTCCCGAGAGACCTCACCCTGAGAGTGCAGACACTTT 129
DB 460 AACCTGATGTGAAGCACTTCTCCCGAGAGACCTCACCCTGAGAGTGCAGACACTTT 519
QY 130 GTGAGATCCACGGAAGACAGAGGCGCGAGGACGACGCTACATTTCCCGTAG 189
DB 520 ATTGAGGTGATGGAAGAAACATGAGAGCGCCAGATGACATGTTTCACTCCAGGAG 579
QY 190 TTCCACCGCCGCTACCGGCTGCGCTGCAACGTGACCAAGTCCCTCTCTTGTCTCCCTG 249
DB 580 TTCACAGGAAGAAATACCGGATCCGAGCTGATGAGACCTCTCACCATTACTTCACTCCTG 639
QY 250 TCTGCCGATGCGATGCTGACCTTCTGTGCGCCCAAGATCCAG 291
DB 640 TCATCTGATGAGGAGTCCCTCACTGTGATGAGACCAAGAAACAG 681

RESULT 14

AAC10867
ID AAC10867 standard; cDNA; 893 BP.

XX AAC10867;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 14942.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

OS

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 14942; 71bp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX mRNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
CC
SQ Sequence 893 BP, 221 A, 259 C, 215 G, 198 T, 0 U, 0 Other;
Query Match 33.2%; Score 123.6; DB 3; Length 893;
Best Local Similarity 64.9%; Pred. No. 8.2e-19;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 10 CGACCGTGTGACTCCGCACTCTGAGGTTGATCCGACCGGACAAAGTTGTATC 69
DB 437 CCCAGCTGTTTGAACACTGACTCTCAGAGATGCGCTGGAGAGACAGGTTCTGTGTC 496
QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGAGACTCAACCGTAGAGTCAGAGACTTT 129
DB 497 AACCTGATGTGAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTTGGAGATGTG 556
QY 130 GTGAGATGCCACGGAAGCAACAGAGCGCCAGAGACCAACGAGTCAATTTCCCGTAG 189
DB 557 ATTGAGGTGATGGAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTTGGAGATGTG 616
QY 190 TTCCACCGCGGCTACCGGCTGCGCTCCAGCTGAGACGATGCGGCTCTTGTCTCCCTG 249
DB 617 TTCACAGGAATACCGGATCCCGAGCTAGATGATAGACCTCTCACCATTTACTTCCCTG 676
QY 250 TTGCGGATGAGCATGATGACTTCTGTGGCCGCCAAGATCCAG 291
DB 677 TCATCTGATGGGTCCTCATCTGTAATGAGCAAGAAACAG 718
RESULT 15
AAC10863
ID AAC10863 standard; cDNA; 911 BP.
XX
AC AAC10863;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 14938.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
XX EPI033401-A2.
XX
XX EPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 14938; 71bp + Sequence Listing; English.
PS

XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
CC
SQ Sequence 911 BP, 221 A, 260 C, 211 G, 219 T, 0 U, 0 Other;
Query Match 33.2%; Score 123.6; DB 3; Length 911;
Best Local Similarity 64.9%; Pred. No. 8.2e-19;
Matches 183; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 10 CGACCGTGTGACTCCGCACTCTGAGGTTGATCCGACCGGACAAAGTTGTATC 69
DB 455 CCCAGCTGTTTGAACACTGACTCTCAGAGATGCGCTGGAGAGACAGGTTCTGTGTC 514
QY 70 TTCTCGATGTGAAGCACTTCTCCCGAGAGACTCAACCGTAGAGTCAGAGACTTT 129
DB 515 AACCTGATGTGAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTTGGAGATGTG 574
QY 130 GTGAGATGCCACGGAAGCAACAGAGCGCCAGAGACCAACGAGTCAATTTCCCGTAG 189
DB 575 ATTGAGGTGATGGAAGCACTTCTCCCGAGAGACTCAAAAGTTAAGTTGGAGATGTG 634
QY 190 TTCCACCGCGGCTACCGGCTGCGCTCCAGCTGAGACGATGCGGCTCTTGTCTCCCTG 249
DB 635 TTCACAGGAATACCGGATCCCGAGCTAGATGATAGACCTCTCACCATTTACTTCCCTG 694
QY 250 TTGCGGATGAGCATGATGACTTCTGTGGCCGCCAAGATCCAG 291
DB 695 TCATCTGATGGGTCCTCATCTGTAATGAGCAAGAAACAG 736
Search completed: September 27, 2004, 12:08:42
Job time : 336 secs

This Page Blank (uspto)